2014

Essays on Industrial Agglomeration and Input-Output Linkages

Zheng Tian
West Virginia University

Essays on Industrial Agglomeration and Input-Output Linkages

Zheng Tian

Dissertation submitted
to the College of Business and Economics
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Economics

Brian Cushing, Ph.D., Chair
Randall Jackson, Ph.D.
Peter Schaeffer, Ph.D.
Donald Lacombe, Ph.D.
Feng Yao, Ph.D.

Department of Economics

Morgantown, West Virginia
2014

Keywords: Industrial agglomeration, input-output linkages, measuring agglomeration, the integrated spatial econometric and input-output model, the spatial pseudo panel industry fixed-effect model, the space-industry filter, the effects estimates, the partitioned effects

Abstract

Essays on Industrial Agglomeration and Input-Output Linkages

Zheng Tian

Industrial agglomeration is an important subject in the field of regional economics. It is an economic phenomenon with two dimensions – space and industry. These two dimensions call for research that can take both spatial and industry factors into consideration regarding industrial agglomeration. Despite the advance of economic theories on agglomeration, empirical studies that satisfactorily incorporate spatial and industrial dimensions are very few. My research aims to enrich the existing methodology for empirical studies by combining spatial econometrics and the input-output model to investigate the effects of input-output linkages on industrial agglomeration.

In this research I propose two methods to integrate spatial econometrics and the input-output model. The first method is a direct application of the embedding strategy in the literature of the integrated econometrics and input-output model, extending it by introducing the spatial intermediate demand variable and estimating the model as if it were a spatial panel data model. The industry dimension replaces the time dimension in the spatial panel data model so that the model in this research is called spatial pseudo panel data model. The second method further explores the spatial pseudo panel data model with a space-industry filter. The space-industry filter enriches the embedding strategy by straightforwardly combining spatial econometrics and the input-output model within a simple model setting. Using the space-industry filter, I define four summary measures of effects estimates that disentangle the impacts of the inter-industry and inter-regional linkages on industrial agglomeration. Based on the summary measures, I explore two ways to partition the effects estimates along the spatial and industry directions. Besides these two methods, I also put forward a bootstrap method to identify the existence of industrial agglomeration using the standardized location quotient.

Acknowledgments

First and foremost, I would like to thank my dissertation chair, Dr. Brian Cushing. He has been always helping me since the very beginning when I got the offer letter from him in the spring of 2008 and decided to come to West Virginia University(WVU). It is my great honor and such a wise decision to have Dr. Cushing as my advisor. Thank you for being so patient with me, giving me very detailed guidance in research, and carefully correcting my grammatical errors in writing.

During the six years in WVU, my greatest achievement is my experience at the Regional Research Institute(RRI). I would like to thank Dr. Randall Jackson, who has acted like my co-advisor even though he does not have such a formal title. Through the ECIO project working with Dr. Jackson, I determined my own research topic. He is more like a friend than my boss at the RRI. Thank you for your help, guidance, support, and trust.

I would like to thank my other dissertation committee members, Dr. Peter Schaeffer, Dr. Donald Lacombe, and Dr. Feng Yao, for their help and advice not only in my dissertation but also in my education and professional development. I also want to thank all professors in the Department of Economics, who have taught me and helped me.

I would also appreciate all fellows at the RRI. I learned a lot from Dr. Mulugeta Kahsai who gave me much advice about my research and career. I want to thank Dr. Christa Court for her proofreading of my papers, thank Juan Tomas Sayago-Gomez for sharing his books and ideas, and thank Caigan McKenzie and Doris Berns for their assistance in my working at the RRI. All of you and Dr. Jackson make me really feel like the RRI is an amicable family.

Finally, I am greatly indebted to my family. I owe the most to my lovely wife, Yuanyuan. She wholeheartedly supports my study by taking care of all chores at home. She gave me the most precious gift, Serena, who brings endless happiness to our family. I thank my mother for educating me on the virtues of honesty, integrity, diligence, and perseverance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Industrial agglomeration is one of the focuses of regional and urban economics. It refers to the geographic concentration of a variety of related industries. Many studies, both theoretical and empirical, have been undertaken in regard to this phenomenon. My dissertation research aims to propose new methods in empirical analysis of industrial agglomeration, with an emphasis on the role of input-output linkages between industries in the formation of agglomeration. The main contribution of my research is integrating spatial econometrics and the input-output model, which are both popular empirical methods in regional economics, in an innovative way.

I design the methodology in this research based on the characteristics of industrial agglomeration. Industrial agglomeration is an economic phenomenon with two dimensions. The first dimension is geographic space. Industries locate proximately in the space, reaping benefits from saving on transport costs, face-to-face communication, sharing knowledge, easy search and match of labor forces, and other factors. The second dimension is industry. The interactions between industries through input-output linkages encourage industries to cluster in a region. Research on industrial agglomeration should address both the space and industry dimensions. My research provides new analytical approaches to studying the effects of industry inter-dependence as well as spatial inter-dependence on industrial agglomeration in one model setting.

Industrial agglomeration is important for regional economic policy makers. It generates employment opportunities, broadens the sources of tax revenue, enhances the investment environment, and stimulates regional economic growth. Industrial agglomeration is a self-reinforcing process in which the inter-dependence among industries through input-output linkages creates cumulative causation, pushing the current level of agglomeration to a higher one. The advantages of industrial agglomeration to a regional economy through the cumulative causation mechanism have been long observed and advocated in the growth-pole strategy by Perroux (1955) and Hanson (1967) and re-polished in the industry cluster strategy by Porter (1990), which guides the practice of regional economic policies. For policy makers, understanding the impacts of input-output linkages on agglomeration is critical to pick up the right target industries that can trigger the self-reinforcing process through inter-industry and spatial linkages, and enhance regional competitiveness.

1.1 A Brief Review of Theories

This section selects some important works in the literature of theories on industrial agglomeration[1], focusing on the relationship between the input-output linkages and agglomeration. I regard these theories as the foundation of my empirical research.

Fujita and Thisse (2002) attribute the earliest theory on agglomeration to von Thünen (1826). Through his practice in agricultural production, von Thünen discovered the critical factors contributing to the formation of a core-periphery pattern between a central town and its surrounding rural area. A large number of classical theories on industrial agglomeration emerged during the early 20th century. Representative works include Central-Place Theory by Christaller (1933) and Lösch (1938), Location Theory by Weber (1909) and Koopmans (1957), Marshallian Agglomeration Economies by Marshall (1890), and the Spatial Competi-

[1] See Fujita et al. (1999), Fujita and Thisse (2002), and Combes et al. (2008) for comprehensive survey of theoretical studies.

tion Model by Hotelling (1929). Among these classical theories, Marshallian Agglomeration Economies are mostly cited, serving as the starting point of many empirical works (Rosenthal and Strange, 2001; Ellison and Glaeser, 1997; Overman and Puga, 2010; Ellison et al., 2010). This concept is often construed to consist of three components, input sharing, labor market pooling, and knowledge spillovers, among which input sharing emphasizes the impact of input-output linkages between industries on agglomeration. In urban economics, the concept of agglomeration economies is often referred to as localization economies and urbanization economies, which are external economies from clustering of firms in a single industry and clustering across industry boundaries (O'Sullivan, 2011; Parr, 2002).

However, in spite of these early valuable contributions, mainstream economics still had trouble explaining agglomeration. At that time when the perfect competition model was the dominating paradigm in economics, economic theories seldom included increasing returns and imperfect competition in the models. Starrett (1978) formally explained the failure of the perfect competition framework in explaining agglomeration, summarized in the Spatial Impossibility Theorem, asserting that if space is homogeneous, the existence of transport costs precludes any perfectly competitive equilibrium.

New Economic Geography (NEG) arising in the 1990s solved the problem of classical theories through the Dixit-Stiglitz monopolistic competition model under the general equilibrium framework. Krugman's (1991) paper initiated NEG by introducing the prototypical Core-Periphery (CP) model. Important contributions to NEG after Krugman (1991) include Krugman and Venables (1995) and Venables (1996) on the input-output linkages, Martin and Ottaviano (1999) and Baldwin and Forslid (2000) on agglomeration and endogenous growth, Fujita and Thisse (2002) on urban structure and system, and Forslid and Ottaviano (2003) and Ottaviano and Robert-Nicoud (2006) on an analytically solvable CP model.

In the NEG theories, formation of agglomeration is the outcome of some fundamental parameters of the economy. These parameters include the degree of increasing returns to scale, the share of consumption in each variety of product, personal income, input-output linkages, and transport costs. Through market mechanisms, specifically the final demand market, the intermediate demand market, and the labor market between regions, these parameters determine the centripetal forces and centrifugal forces of agglomeration. Spatial equilibrium is either an agglomeration in one region or a symmetric distribution between two regions, depending on the interaction between the two opposing forces.

Input-output linkages between industries play an important role in agglomeration in the NEG theories. The Vertical Linkage (VL) model advanced by Krugman and Venables (1995) explains the impacts of input-output linkages on international division of manufacturing production. The VL model was further improved by Venables (1996), Puga and Venables (1996), Ottaviano and Robert-Nicoud (2006), Kranich (2011), to fit into other situations. Input-output linkages generate forward linkages and backward linkages that pull industries into a region. The forward linkages come from the lower prices of intermediate inputs into the upstream industries due to saving on transport costs when industries locate closer. The backward linkages come from intermediate demand in the downstream industries and consumption demand for the final goods. The forward and backward linkages yield strong centripetal forces for industrial agglomeration.

In sum, according to both classical and NEG theories, industrial agglomeration results from the following basic ingredients: increasing returns to scale, final demand market, input-output linkages or intermediate demand market, and transport costs. The spatial equilibrium is the outcome of interactions of these factors across regions and industries, implying a non-linear cumulative causation process of spatial distribution. I summarize these ideas using equation (1.1), which can be considered as a reduced-form solution for the VL model. For an economy with N regions and S industries, λ_{ir} is the share of production of industry i in region r in the overall economy. Then, λ_{ir} is determined by the following equation

$$\lambda_{ir} = f(\lambda_{js}, \tau_{rs}, \alpha_{ij}, \mu_{ir}, \sigma_{ir}) \text{ for } i, j = 1, \ldots, N, \text{ and } r, s = 1, \ldots, S \tag{1.1}$$

The basic idea of equation (1.1) is that λ_{ir} is determined by λ_{js} and the fundamental parameters in all regions and industries. τ_{rs} is the transport costs between regions r and s, α_{ij} is the input-output relationship between industries i and j, μ_{ir} is the share of consumption of products of industry i in personal income of region r, and σ_{ir} represents the degree of increasing returns to scale of industry i in region r. The goal of my research is to empirically examine equation (1.1), exploring new methods to modeling the inter-dependence of both regions and industries implied in the equation.

1.2 The Overview of the Dissertation

A prerequisite for studying industrial agglomeration is to find a measure to indicate the existence and degree of industrial agglomeration in a region. Measuring agglomeration is an important aspect of empirical research in regional economics. Among a variety of agglomeration indices, in Chapter 2, I choose the location quotient (LQ) to measure agglomeration. The LQ is widely used in regional economics due to its simplicity in computation. A LQ greater than one for an industry in a region is an indication of agglomeration. However, researchers may also choose their specific cut-off value to define an agglomeration, such as a LQ greater than 1.5 or 2. The arbitrary delimitation makes the usefulness of the LQ as an agglomeration index questionable. To obtain an objective cut-off value, I follow the idea in the paper of O'Donoghue and Gleave (2004) who propose a standardized location quotient (SLQ), and use the 5% critical value of the standard normal distribution as the cut-off value for the SLQ. In Chapter 2, I suggest a bootstrap method to determine the cut-off value for the SLQ. The advantage of the new method is that it does not rely on any assumption regarding the statistical distribution of the LQ, which is a major limitation of O'Donoghue and Gleave's (2004) approach. And I apply the bootstrap method to measure agglomeration of manufacturing industries at the county level in the United States.

The study in Chapter 3 is an attempt to empirically test equation (1.1), focusing on the role of input-output linkages in agglomeration. The methodology used in this chapter is to extend the embedding strategy in the literature of the integrated econometric and input-output model (Rey, 2000) through introducing spatial econometrics into the model. Including spatial econometrics is critical to account for spatial spillover effects in industrial agglomeration, while the input-output model is capable of modeling input-output linkages between industries. The integrated spatial econometric and input-output model addresses the two-dimensional characteristic of industrial agglomeration mentioned above. The integration technique is to create the spatial intermediate demand variable (SIDV) using the input-output matrix and the spatial weight matrix, and then embed this variable in a spatial econometric model. The SIDV can be computed separately as the local intermediate demand variable (IDV) and its spatial lag so that the spatial econometric model can be naturally a spatial Durbin model. With the LQ as the dependent variable, I regress it on explanatory variables, including the intermediate variable, market potential, and other control variables, as well as their spatial lags. The nature of the dataset and the construction of the key variables suggest estimating the spatial Durbin model by "borrowing" spatial panel data model techniques. Unlike a panel data model in the normal sense, the time dimension in the panel is replaced with the industry dimension, meaning that for one region the sample has observations for several industries instead of several time periods. I estimate the model with the maximum likelihood method used in the spatial panel data model with time fixed effects. In my model industry fixed effects substitute for time fixed effects. I refer to this model as a spatial pseudo-panel industry fixed effects Durbin model. I run the model using the sample containing all manufacturing industries with 3-digit NAICS codes at the county level in the United States. The results confirm the importance of input-output linkages for industrial agglomeration in a region.

In Chapter 4 I construct a space-industry filter to explore a new way of combining spatial econometrics and the input-output model. The space-industry filter originates from the space-time filter in Parent and LeSage (2012), replacing the time component in the space-time filter, which is a matrix composed of time lag operators, with an industry component using an input-output matrix. Unlike the embedding strategy that creates an independent variable using an input-output matrix and then embeds it into the model, the space-industry filter works on the dependent variable, straightforwardly modeling the inter-regional and inter-industry dependence in this variable. After imposing the space-industry filter on the dependent variable, a linear model is transformed into a spatial econometric model, with the filter working as a composite weight matrix. This new approach successfully models the general equilibrium of the spatial distribution of industries, as indicated in equation (1.1). The model is estimated using the Bayesian MCMC method, with a Monte Carlo experiment guaranteeing its validity. I also define the effects estimates for the model. The space-industry filter introduces non-linearity into the model so that we cannot directly interpret the coefficients on independent variables as their partial derivatives with respect to the dependent variable. There are four summary measures of effects estimates, the total space-industry effects, the direct space-industry effects, the direct within industry effects, and the indirect space-industry effects. On one side, these effects estimates synthesize the multiplier effects from the input-output model and the spatial spillover effects from spatial econometrics. On the other side, these effects estimates decompose the total inter-dependence across regions and industries into the inter-regional dependence and the inter-industry dependence. The summary measures can be further partitioned along spatial and industry dimensions. I apply the space-industry filter approach to estimate a simple regional industry employment model at the state and county levels in the United States with results showing strong positive effects of input-output linkages on industry distribution in space.

Chapter 5 concludes the dissertation and sets forth some directions for future research. The new methods proposed in this research, especially the space-industry filter method, are preliminary but promising. I will envision how I can apply the space-industry filter method to a wide range of topics.

Chapter 2

Measuring Agglomeration Using the Standardized Location Quotient with a Bootstrap Method

2.1 Introduction

Constructing an index to measure industrial agglomeration is an important aspect of empirical studies in regional economics. Economists have long been seeking to develop an index that can accurately reflect the degree of agglomeration across industries, time, and space. Among the indices that have been examined, this paper focuses on the location quotient (LQ)[1], which is used widely in regional science due to the simplicity in computation and low data requirements.

The LQ measures the ratio between the local and national share of productive activities of a particular industry in a region. A LQ greater than one indicates that the industry under study is more concentrated in the region than the national average. Apart from using one

[1]The LQ is often referred to as an index of localization of an industry. Following the terminology in the paper by Nakamura and Paul (2009), I consider the LQ as an index of agglomeration, treating agglomeration and localization as synonyms.

as the cut-off value, some researchers use values such as 1.25 or 2 to detect the presence of agglomeration in a region. The subjective determination of the cut-off value can raise doubts about how large the value of the LQ should be to ensure the existence of agglomeration.

To find an objective cut-off value of the LQ for identifying agglomeration, O'Donoghue and Gleave (2004) put forward an approach of computing the standardized location quotient (SLQ), which is simply the z-statistics of the LQ, and using the 5% critical value of the standard normal distribution as the cut-off value of the SLQ, given that the LQ is normally distributed. The limitation of the SLQ approach is that if the LQ does not follow the normal distribution, then the cut-off value determined by this approach is not reliable. Therefore, following the idea of the SLQ approach, I suggest an alternative method to obtain an objective cut-off value of the LQ without any assumption regarding the statistical distribution. Then this method is applied to measure agglomeration of manufacturing industries in U.S counties.

2.2 Literature Review

Combes et al. (2008, Chapter 10) and Nakamura and Paul (2009) provide a comprehensive literature review that covers most of the existing indices. Rather than duplicate their work,I will briefly survey some important aspects about the existing indices that motivate modifying the SLQ method to measure industrial agglomeration.

The existing indices can be categorized into two types: discrete and continuous. The discrete indices apply to areal data on discrete spatial units, such as counties, states, and countries. The majority of agglomeration indices belong to the discrete type, including the LQ, the Gini index, the Theil index, the Isard index, the Herfindahl-Hirschman (HH) index, the Ellison-Glaeser (EG) index (Ellison and Glaeser, 1997), and the Maurel-Sedillot (MS) index (Maurel and Sédillot, 1999), a variant of the EG index. Among the discrete indices, the EG index has been widely adopted by researchers in measuring industrial agglomeration

(Rosenthal and Strange, 2001; Holmes and Stevens, 2004; Bertinelli and Decrop, 2005). The dartboard framework under which the EG index is derived also initiates a way to discover other indices and to introduce statistical tests on indices. (Maurel and Sédillot, 1999; Guimarães et al., 2009) However, the EG and MS index, as well as the Gini index, suffer a problem in their application. They are unable to evaluate the degree of agglomeration of an industry in a particular region because the regional dimension is integrated out in the process of computation. Another problem that all discrete indices encounter is the modifiable areal unit problem (MAUP). The problem arises when the artificial change in the delimitation of discrete spatial units significantly alters the reported distribution of economic activities among the spatial units, leading to a biased measure of agglomeration. The continuous indices are designed to control for the MAUP problem.

With the assumption of a continuous space, the continuous indices are applied to spatial point objects which are represented by geographical coordinates. A typical continuous index measures the density of economic activities along the distance between each pair of points. The studies exploring the continuous-type indices include the Duranton-Overman K-density in Duranton and Overman (2005) and Ripley's K-function used by Marcon and Puech (2003, 2010) and Arbia et al. (2010). However, a practical problem of the continuous indices is that the data requirements are so high that ordinary researchers without access to datasets with detailed information are unable to compute a continuous index. Moreover, the continuous indices usually measure agglomeration without reference to any administrative entities so that their implications for local economic policy makers are not readily applicable.

Establishing a guideline for creating new agglomeration indices, Duranton and Overman (2005) bring forth five properties that an index should satisfy. Combes et al. (2008, Chapter 10) put forward another three properties. Together, these properties include that an agglomeration index should be (1) be comparable across industries, (2) be comparable across spatial scales, (3) be unbiased with respect to arbitrary changes to spatial classification, (4) be unbiased with respect to arbitrary changes to industrial classification, (5) control for overall distribution of economic activities, (6) allow for a significance test, (7) be computable in the closed-form from accessible data, and (8) be justified by a suitable model.

Although it is desirable to find an agglomeration index that can meet most of the properties, the index that a researcher actually chooses is often constrained by the data availability and the purpose of the study. The purpose of this paper is to identify industrial agglomeration in an administrative spatial unit, implying that neither the Gini and EG indices nor the continuous indices are suitable for this purpose as the former provide no information regarding agglomeration in a specific spatial unit and the latter are applied only to spatial point objects without clear reference to an administrative unit. In contrast, the LQ-type index suffices to serve the purpose of this paper, taking advantage of its flexibility in application at any level of both industrial and geographical aggregation and the ease of collecting data for states and counties in the U.S.

As pointed out in the introduction section, a problem of using the LQ is concerned with how to objectively determine the cut-off value for defining agglomeration. Choosing any arbitrary value of the LQ as the cut-off value can always call the validity into question. Besides the SLQ approach in O'Donoghue and Gleave (2004), two other studies attempt to address this problem by building some statistical tests to determine the cut-off value. Moineddin et al. (2003) derive an expression of the standard deviation of the LQ and then construct the confidence interval by assuming the normal distribution of the LQ. Following the dartboard framework of Ellison and Glaeser (1997), Guimarães et al. (2009) provide a theoretical foundation of the LQ and then derive two test statistics that are asymptotically chi-squared distributed. One common drawback of the three studies is that assumptions made on the statistical distribution of the LQ and test statistics may not hold so that their results are still questionable. To overcome this drawback, I propose a method that does not depend on any assumption regarding the statistical distribution. I choose to extend the idea of the SLQ approach because it is not only easy to implement but also subjected to fewer assumptions than in the other two studies.

2.3 The Standardized Location Quotient

Let $i = 1, 2, \ldots, I$ denote industries and $j = 1, 2, \ldots, J$ denote regions, then the LQ of industry i in region j is defined as

$$LQ_{ij} = \frac{s_{ij}}{s_{*j}} = \frac{x_{ij}/x_{i*}}{x_{*j}/x_{**}} \tag{2.1}$$

where x_{ij} represents employment of industry i in region j, x_{i*} is total employment of industry i in all regions, x_{*j} is total employment of all industries in region j, and x_{**} is total employment of the overall economy. Thus, s_{ij} is the share of industry i's employment in region j relative to total employment of industry i, and s_{*j} is the share of region j's employment relative to total employment in the overall economy. If $LQ_{ij} > 1$, then industry i is said to be concentrated in region j.

There are two problems in the statistical properties of the LQ. First, the LQ is notoriously skewed in that the minimum value could be zero when industry i has no production in region j and the maximum value could reach as high as two or three hundred when industry i is very concentrated in region j. The median and mean values are around one to three. The skewness of the LQ implies that there could be influential observations that affect the estimation of coefficients in the regression model. Second, the cut-off value of the LQ that determines whether industry i is concentrated in region j is often arbitrarily determined. Thus, a relatively objective way to obtain the cut-off value is needed.

Arguing against using unity or other arbitrary values of the LQ to define agglomeration, O'Donoghue and Gleave (2004) propose using the standardized location quotient, simply the z-statistic of the original LQ. The SLQ of industry i in region j is,

$$SLQ_{ij} = \frac{LQ_{ij} - \overline{LQ_i}}{std(LQ_i)} \tag{2.2}$$

where $\overline{LQ_i}$ and $std(LQ_i)$ are the mean and standard deviation of the LQ of industry i.

Before being converted to the z-statistic, the LQ is tested to see whether it is normally distributed using the Kolmogorov-Smirnov normality test. If the test fails to confirm normality, the logarithmic function is used to transform the LQ, followed by a test of the normality of $\log(LQ)$. Passing the normality test implies that the SLQ, or the standardized $\log(LQ)$, should conform to the standard normal distribution. The cut-off level for confirming the existence of agglomeration in a region is then determined by the critical value of the standard normal distribution at the 5% level, i.e., 1.96 for a two-tailed test or 1.64 for a one-tailed test.

The crucial point in O'Donoghue and Gleave's (2004) approach hinges on the assumption that the LQ is normally distributed. Hence, the most serious limitation of the approach is that the critical value of the normal distribution may not be reliable if the normality assumption is invalid. O'Donoghue and Gleave (2004) acknowledge this limitation and suggest not using the SLQ approach if the normality test on the LQ fails, but the authors provide no alternatives to solve the problem. Moreover, an implicit assumption in this approach is that the statistical distributions of the LQ indices of all industries are the same. This assumption is questionable because the actual data generating process of the LQ indices may be different among industries, determined by some industry-specific characteristics.

A simple regression model can illustrate the problem more clearly. Consider the following model

$$\boldsymbol{LQ_i} = \alpha\, \boldsymbol{\iota} + \boldsymbol{u_i} \tag{2.3}$$

where $\boldsymbol{LQ_i}$ is a $J \times 1$ vector of the LQ of industry i, and $\boldsymbol{\iota}$ is a $J \times 1$ vector of 1's, and $\boldsymbol{u_i}$ is a vector of random errors following a certain statistical distribution. It follows that the SLQ is simply the residuals from the ordinary least square estimation of Equation (2.3), divided by the standard deviation of the residuals. If, in the actual data generating process, $\boldsymbol{u_i}$ is not normally distributed, then neither is the distribution of the residuals, i.e., the SLQ. Further, it is not necessarily true that $\boldsymbol{u_i}$'s for all $i = 1, 2, \ldots, I$ have the same distribution. Therefore, there is no well-founded justification for merely using the critical value at the

5% level of the standard normal distribution to determine the cut-off level of the LQ for all industries.

To circumvent the obstacle incurred by the normality assumption, I propose using the bootstrap method to get the estimated 5%-level critical value of the actual distribution of the SLQ. Essentially, the bootstrap method is based on the Fundamental Theorem of Statistics asserting that the empirical distribution function of a random variable X consistently estimates the true cumulative distribution function of X. It follows that test statistics constructed from the empirical distribution are also the consistent estimators of the exact statistics from the true distribution (Davidson and MacKinnon, 2004).

The steps of using the bootstrap method to determine the cut-off level are as follows,

1. computing the SLQ for each industry in all regions;

2. performing bootstrap resampling of the SLQ for each industry. A bootstrap resampling is a process of randomly drawing samples from the whole original sample with replacement. A bootstrapped sample has the same length as the original sample;

3. obtaining the 95^{th} percentile for each bootstrapped sample of an industry and performing the bootstrap sampling N (set $N = 999$) times to get a set of N 95^{th} percentiles. The purpose of this step is to draw the 95^{th} percentile from its empirical distribution;

4. using the sample mean of the bootstrap 95^{th} percentiles as the estimate of the critical value at the 5% level of the true distribution.

Using the bootstrap method, the determination of the cut-off value does not rely on any assumption of the statistical distribution. Moreover, the cut-off value of the SLQ for each industry is unique since the bootstrap sampling is carried to individual industries independently. The most notable advantage of using the bootstrap method is the simplicity in implementation. This method enhances the applicability of the original SLQ approach by relaxing the normality assumption.

In this study, I do not intend to address all issues regarding an agglomeration index to satisfy the properties proposed by Duranton and Overman (2005) and others. However, the bootstrap method can be construed as partly solving the sixth property requiring a test on the statistical significance of an index. To meet other properties, a more theoretically founded method is needed to derive an agglomeration index, which is definitely not an easy task but a promising direction for future studies.

2.4 An Application to Manufacturing Industries in U.S. Counties

To evaluate the effectiveness of the bootstrap method to define the cut-off value, I apply this method to manufacturing industries in the United States.

2.4.1 Data

The data source is the 2002 County Business Patterns (CBP) imputed by Isserman and Westervelt (2006). The CBP, published by the U.S. Bureau of the Census, contains a comprehensive annual compilation of information about the location of establishments with employment in the United States. The CBP data set is constructed with a hierarchical structure, in which industries are categorized by the two to six digit NAICS codes and spatial units cover four levels of spatial aggregation: the nation, states, counties, and zip-code areas. However, the nondisclosure problem of the CBP impairs the usefulness of the data set. For the purpose of protecting private business confidentiality, some data on employees are suppressed by the Bureau of the Census. Instead, employment flags are used to indicate the range of the missing value of an industry/county pair with undisclosed data.

Isserman and Westervelt (2006) propose a two-stage method to complete the CBP data set. The first stage identifies the smallest possible range for each withheld data value given the information provided by employment flags. Taking advantage of the hierarchical structure of the CBP data set, the second stage estimates the missing values and iteratively adjusts each estimate to ensure that the estimated number of employees adds up correctly to the total number of employees in the higher levels of aggregation along both the industrial and spatial hierarchies. I downloaded their complete data set of the 2002 CBP from `www.wholedata.com`[2].

In this paper I use the county-level employment of manufacturing industries with 3-digit NAICS codes to compute the SLQ. Table. 2.1 shows the names of all manufacturing industries with 3-digit NAICS codes. The total sample size is 35,328 with each observation being a unique industry/county pair, which in total consists of 21 manufacturing industries and 3,076 counties.

Table 2.1: The 3-digit-NAICS Manufacturing Industries

	NAICS	NAMES
1	311	Food mfg
2	312	Beverage & tobacco product mfg
3	313	Textile mills
4	314	Textile product mills
5	315	Apparel manufacturing
6	316	Leather & allied product mfg
7	321	Wood product mfg
8	322	Paper mfg
9	323	Printing & related support activities
10	324	Petroleum & coal products mfg
11	325	Chemical mfg
12	326	Plastics & rubber products mfg
13	327	Nonmetallic mineral product mfg
14	331	Primary metal mfg
15	332	Fabricated metal product mfg
16	333	Machinery mfg
17	334	Computer & electronic product mfg
18	335	Electrical equip, appliance & component mfg
19	336	Transportation equipment mfg
20	337	Furniture & related product mfg
21	339	Miscellaneous mfg

[2]Unfortunately, this website is currently shut down.

2.4.2 Results

The LQ indices are computed in four forms: LQ, $\log(LQ)$, SLQ, and standardized $\log(LQ)$ (referred to as SLLQ hereafter). In this section, I first report the effects of standardization and the logarithmic transformation on the skewness problem and the statistical distribution of the original LQ indices. Then I use the bootstrap method to obtain the cut-off values of the SLQ and SLLQ for all manufacturing industries.

Table 2.2 shows that the LQ indices for all manufacturing industries exhibit serious skewness. The gap between the median and mean is considerable, and the ranges of the LQ indices differ widely among various industries. The minimum values of the LQ indices for all manufacturing industries are approximately zero, but the maximum values can be as high as 338.20 (Beverage and Tobacco Product Manufacturing) and as low as 27.05 (Computer and Electronic Product Manufacturing). Also, only a few values of the means of the LQ indices are close to unity (Printing and Related Support Activities and Computer and Electronic Product Manufacturing), and the variation of the means of all industries is also substantial. This invalidates the practice of using unity or any other arbitrary value as the cut-off value for identifying agglomeration.

The logarithmic transformation and standardization operation on the LQ can effectively alleviate the skewness problem. As shown in Table 2.3, the gap between the means and medians of $\log(LQ)$ values is remarkably narrowed. For example, the median and mean in Food Manufacturing are almost equal. Moreover, the difference in the maximum values of the $\log(LQ)$ indices between industries are not as pronounced as the LQ indices, and the range between the minimum and maximum values of $\log(LQ)$ is, to some extent, balanced around the means for most industries. Compared with the logarithmic transformation, standardization can also reduce the gap between the means and medians, but the range of the SLQ values is not as balanced as $\log(LQ)$. (see Table 2.4) Finally, the SLLQ combines the effects of both the logarithmic transformation and standardization, resulting a balanced range around the mean and median that are close to zero. (see Table 2.5)

Table 2.2: The Summary Statistics of LQ of Manufacturing Industries

	Min.	Median	Mean	Max.
Food mfg	0.00	0.63	2.50	65.54
Beverage & tobacco product mfg	0.00	0.57	2.82	338.20
Textile mills	0.00	0.43	7.40	170.90
Textile product mills	0.01	0.45	4.16	250.10
Apparel manufacturing	0.00	0.43	3.76	134.10
Leather & allied product mfg	0.01	0.88	8.30	298.70
Wood product mfg	0.00	1.45	4.78	81.22
Paper mfg	0.00	1.23	4.28	85.68
Printing & related support activities	0.01	0.39	1.09	47.93
Petroleum & coal products mfg	0.00	0.44	3.96	193.20
Chemical mfg	0.00	0.53	1.89	71.64
Plastics & rubber products mfg	0.00	1.01	2.41	55.25
Nonmetallic mineral product mfg	0.01	0.82	2.11	82.63
Primary metal mfg	0.00	0.79	3.43	106.70
Fabricated metal product mfg	0.00	0.79	1.46	34.06
Machinery mfg	0.00	0.79	1.86	65.28
Computer & electronic product mfg	0.00	0.33	1.13	27.05
Electrical equip, appliance & component mfg	0.00	0.86	3.36	87.68
Transportation equipment mfg	0.00	0.63	1.93	47.27
Furniture & related product mfg	0.01	0.50	2.39	113.50
Miscellaneous mfg	0.01	0.45	1.44	52.42

Table 2.3: The Summary Statistics of log(LQ) of Manufacturing Industries

	Min.	Median	Mean	Max.
Food mfg	-5.57	-0.46	-0.46	4.18
Beverage & tobacco product mfg	-5.81	-0.56	-0.61	5.82
Textile mills	-6.15	-0.85	-0.63	5.14
Textile product mills	-4.91	-0.79	-0.56	5.52
Apparel manufacturing	-5.99	-0.84	-0.66	4.90
Leather & allied product mfg	-5.13	-0.13	0.00	5.70
Wood product mfg	-5.94	0.37	0.29	4.40
Paper mfg	-6.00	0.20	0.11	4.45
Printing & related support activities	-4.48	-0.95	-0.88	3.87
Petroleum & coal products mfg	-5.53	-0.83	-0.57	5.26
Chemical mfg	-5.84	-0.63	-0.72	4.27
Plastics & rubber products mfg	-6.77	0.01	-0.18	4.01
Nonmetallic mineral product mfg	-5.01	-0.20	-0.14	4.41
Primary metal mfg	-6.57	-0.23	-0.32	4.67
Fabricated metal product mfg	-6.68	-0.23	-0.40	3.53
Machinery mfg	-5.50	-0.23	-0.38	4.18
Computer & electronic product mfg	-6.45	-1.10	-1.23	3.30
Electrical equip, appliance & component mfg	-5.99	-0.15	-0.32	4.47
Transportation equipment mfg	-6.92	-0.46	-0.74	3.86
Furniture & related product mfg	-4.90	-0.70	-0.59	4.73
Miscellaneous mfg	-4.90	-0.79	-0.79	3.96

Table 2.4: The Summary Statistics of Standardized LQ of Manufacturing Industries

	Min.	Median	Mean	Max.
Food mfg	-0.47	-0.35	-0.00	11.77
Beverage & tobacco product mfg	-0.23	-0.18	-0.00	26.96
Textile mills	-0.42	-0.40	0.00	9.30
Textile product mills	-0.28	-0.25	0.00	16.48
Apparel manufacturing	-0.38	-0.34	-0.00	13.24
Leather & allied product mfg	-0.32	-0.29	0.00	11.24
Wood product mfg	-0.56	-0.39	-0.00	8.95
Paper mfg	-0.45	-0.32	-0.00	8.55
Printing & related support activities	-0.39	-0.26	0.00	17.05
Petroleum & coal products mfg	-0.28	-0.25	0.00	13.52
Chemical mfg	-0.39	-0.28	-0.00	14.42
Plastics & rubber products mfg	-0.60	-0.35	0.00	13.18
Nonmetallic mineral product mfg	-0.47	-0.29	0.00	17.92
Primary metal mfg	-0.42	-0.33	0.00	12.78
Fabricated metal product mfg	-0.67	-0.31	-0.00	14.93
Machinery mfg	-0.57	-0.33	-0.00	19.31
Computer & electronic product mfg	-0.50	-0.35	0.00	11.50
Electrical equip, appliance & component mfg	-0.46	-0.34	-0.00	11.55
Transportation equipment mfg	-0.54	-0.36	0.00	12.70
Furniture & related product mfg	-0.32	-0.25	-0.00	14.89
Miscellaneous mfg	-0.44	-0.30	-0.00	15.73

Table 2.5: The Summary Statistics of Standardized log(LQ) of Manufacturing Industries

	Min.	Median	Mean	Max.
Food mfg	-2.90	-0.00	-0.00	2.63
Beverage & tobacco product mfg	-2.90	0.03	-0.00	3.59
Textile mills	-2.12	-0.09	-0.00	2.21
Textile product mills	-2.33	-0.12	0.00	3.25
Apparel manufacturing	-2.46	-0.08	0.00	2.57
Leather & allied product mfg	-2.42	-0.06	-0.00	2.68
Wood product mfg	-3.51	0.05	0.00	2.32
Paper mfg	-3.40	0.05	0.00	2.42
Printing & related support activities	-2.70	-0.05	-0.00	3.57
Petroleum & coal products mfg	-2.68	-0.14	-0.00	3.15
Chemical mfg	-2.96	0.05	0.00	2.89
Plastics & rubber products mfg	-3.90	0.11	-0.00	2.48
Nonmetallic mineral product mfg	-3.76	-0.05	0.00	3.51
Primary metal mfg	-3.12	0.04	-0.00	2.49
Fabricated metal product mfg	-4.59	0.12	0.00	2.87
Machinery mfg	-3.24	0.09	0.00	2.88
Computer & electronic product mfg	-2.79	0.07	0.00	2.42
Electrical equip, appliance & component mfg	-2.83	0.09	0.00	2.40
Transportation equipment mfg	-3.15	0.14	0.00	2.34
Furniture & related product mfg	-2.70	-0.07	-0.00	3.34
Miscellaneous mfg	-2.68	-0.00	-0.00	3.10

The normality test of the SLQ and SLLQ provides little support for the normality assumption, which is the key in ODonoghue and Gleaves (2004) approach. The SLQ indices for all industries fail to pass the Kolmogorov-Smirnov normality test. Table 2.6 shows that the p-values for all SLQ indices are approximately zero, which means the null hypothesis of the normal distribution is rejected. Although the test result for the SLLQ is better than for the SLQ, the SLLQ indices for only three industries (Food Manufacturing, Beverage and Tobacco Product Manufacturing, and Miscellaneous Manufacturing) have p-values greater than 0.05, supporting the normality assumption. Figure 2.1 displays histograms of the SLQ and SLLQ for Food Manufacturing. The histogram of the SLLQ looks more like a normal distribution than does that of the SLQ.

Table 2.6: The Normality Test of the SLQ and SLLQ Indices

	SLQ		SLLQ	
	statistic	p.value	statistic	p.value
Food mfg	0.32	0.00	0.01	0.41
Beverage & tobacco product mfg	0.41	0.00	0.02	0.24
Textile mills	0.34	0.00	0.07	0.00
Textile product mills	0.39	0.00	0.06	0.00
Apparel manufacturing	0.35	0.00	0.04	0.00
Leather & allied product mfg	0.37	0.00	0.04	0.01
Wood product mfg	0.29	0.00	0.03	0.00
Paper mfg	0.33	0.00	0.06	0.00
Printing & related support activities	0.35	0.00	0.02	0.04
Petroleum & coal products mfg	0.39	0.00	0.06	0.00
Chemical mfg	0.35	0.00	0.02	0.01
Plastics & rubber products mfg	0.27	0.00	0.05	0.00
Nonmetallic mineral product mfg	0.32	0.00	0.03	0.00
Primary metal mfg	0.34	0.00	0.04	0.00
Fabricated metal product mfg	0.25	0.00	0.05	0.00
Machinery mfg	0.29	0.00	0.04	0.00
Computer & electronic product mfg	0.31	0.00	0.05	0.00
Electrical equip, appliance & component mfg	0.32	0.00	0.04	0.00
Transportation equipment mfg	0.29	0.00	0.06	0.00
Furniture & related product mfg	0.37	0.00	0.05	0.00
Miscellaneous mfg	0.33	0.00	0.02	0.18

The failure of the normality test for the LQ indices for most industries suggests using the

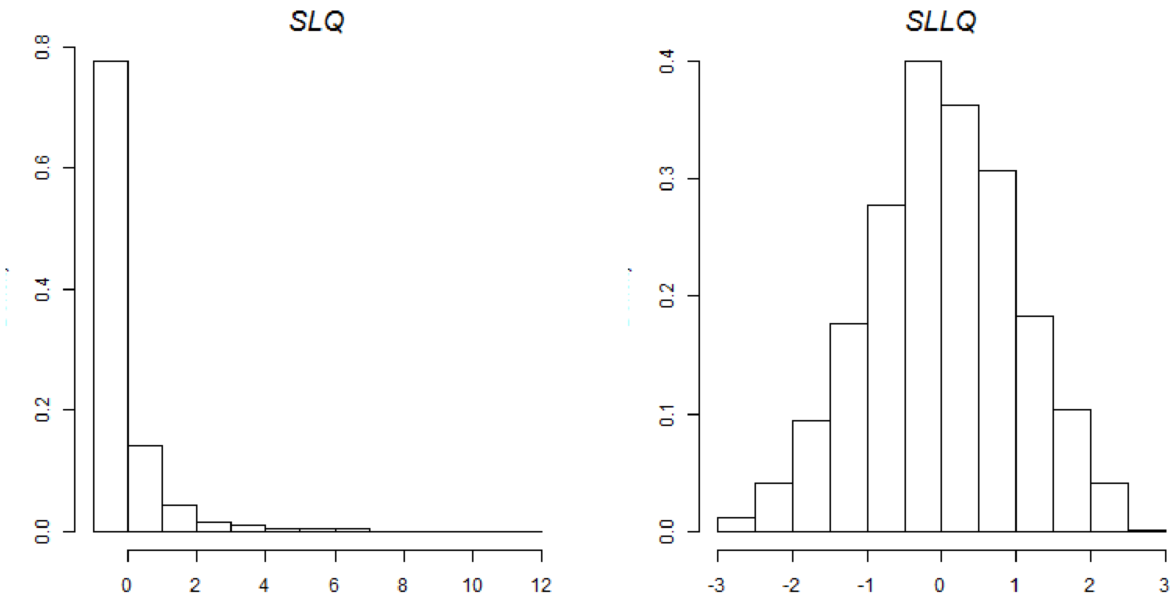

Figure 2.1: Comparison of Histograms Between SLQ and SLLQ of Food Mfg

bootstrap method to obtain the cut-off value to delimit agglomeration. With the bootstrap method, Tables 2.7 and 2.8 present the cut-off values of manufacturing industries based on the bootstrapped SLQ and SLLQ, respectively. As shown in the first column in both tables, the cut-off values for most industries are different from 1.64, the 5% critical value of the standard normal distribution for a one-tailed test. Six industries have cut-off values greater than 1.64 using the SLQ, while nine industries do using the SLLQ. If 1.64 is used as the cut-off value, as O'Donoghue and Gleave (2004) suggest, for more than half of 3-digit NAICS manufacturing industries, we would identify fewer counties as having agglomeration of a particular industry than if we use the bootstrap method to define the cut-off value.

The cut-off values vary across industries, which is expected as the bootstrap method is applied to each industry independently. With the SLQ indices, the lowest cut-off level, 0.78, is for Beverage and Tobacco Product Manufacturing and the highest level, 2.01, for Wood Product Manufacturing. The cut-off values of the SLLQ indices fall into a narrower interval than the SLQ indices, ranging from the minimum of 1.43 to the maximum of 1.95. This is because the logarithmic transformation diminishes the leverage effect from the large values.

Table 2.7: The Cut-off Value of the SLQ with the Bootstrap Method

	The Cut-off Value	Ratio of Employment	Number of Counties
Food mfg	1.63	0.14	125
Beverage & tobacco	0.78	0.27	51
Textile mills	1.98	0.19	44
Textile product mills	1.17	0.30	73
Apparel manufacturing	1.50	0.11	74
Leather & allied product	1.14	0.21	31
Wood product	2.01	0.13	121
Paper mfg	1.56	0.12	58
Printing & related support activities	0.94	0.12	108
Petroleum & coal products	1.05	0.27	43
Chemical mfg	1.15	0.16	85
Plastics & rubber	1.70	0.10	91
Nonmetallic mineral	1.31	0.12	125
Primary metal	1.36	0.20	69
Fabricated metal	1.68	0.08	131
Machinery mfg	1.53	0.09	112
Computer & electronic	1.64	0.26	68
Electrical equip mfg	1.66	0.12	60
Transportation equipment	1.86	0.16	93
Furniture & related product	0.94	0.27	107
Miscellaneous mfg	1.49	0.12	104

Table 2.8: The Cut-off Level of the SLLQ with the Bootstrap Method

	The Cut-off Value	Ratio of Employment	Number of Counties
Food mfg	1.63	0.14	125
Beverage & tobacco	1.75	0.27	51
Textile mills	1.67	0.19	44
Textile product mills	1.95	0.30	73
Apparel manufacturing	1.65	0.11	74
Leather & allied product	1.70	0.21	31
Wood product	1.58	0.14	123
Paper mfg	1.58	0.11	57
Printing & related support activities	1.64	0.12	108
Petroleum & coal products	1.88	0.27	44
Chemical mfg	1.58	0.16	86
Plastics & rubber	1.43	0.10	91
Nonmetallic mineral	1.70	0.13	126
Primary metal	1.49	0.20	70
Fabricated metal	1.48	0.08	131
Machinery mfg	1.46	0.09	112
Computer & electronic	1.50	0.26	68
Electrical equip mfg	1.53	0.12	60
Transportation equipment	1.47	0.16	93
Furniture & related product	1.77	0.27	108
Miscellaneous mfg	1.71	0.12	104

Interestingly, the average of the cut-off values of the SLLQ is 1.63, close to the 5% critical value of the standardized normal distribution, which means that the cut-off value defined by O'Donoghue and Gleave (2004) is only reasonable in an average sense, but it is not accurate for each individual industry.

The second and third columns in Tables 2.7 and 2.8 show that, for counties that are identified as having agglomeration of a particular industry, the shares of employment that these counties account for are nearly identical for using both the SLQ and SLLQ. Also, there is only marginal discrepancy in the number of these counties between the two tables. We cannot get the same kind of results if we rely on the standard normal distribution to get the cut-off values because the SLQ and SLLQ have different statistical distributions. In contrast, the bootstrap method estimates the cut-off values, the means of the bootstrap samples of the 95th percentiles, from the empirical distribution that is the consistent estimator of the true data generating process, regardless of statistical distributions of the SLQ and SLLQ. Therefore, we can identify the same set of counties as having agglomeration of some industry using either the SLQ or the SLLQ even though they have different distributions.

Table 2.7 and 2.8 also illustrate that the degree of agglomeration is disparate across manufacturing industries. For example, 8% employment of Fabricated Metal Manufacturing is distributed in 131 counties. In contrast, in Leather and Allied Product Manufacturing, only 31 counties account for over 20% of industrial employment. Mapping makes such comparison more clear. Figure 2.2 and 2.3 show that Fabricated Metal Manufacturing is more widely spread across counties than Leather and Allied Product Manufacturing, which is concentrated in only a few counties.

2.5 Conclusion

This paper provides a simple bootstrap method to obtain the cut-off value based on the SLQ for identifying the existence of industrial agglomeration in a region. The bootstrap

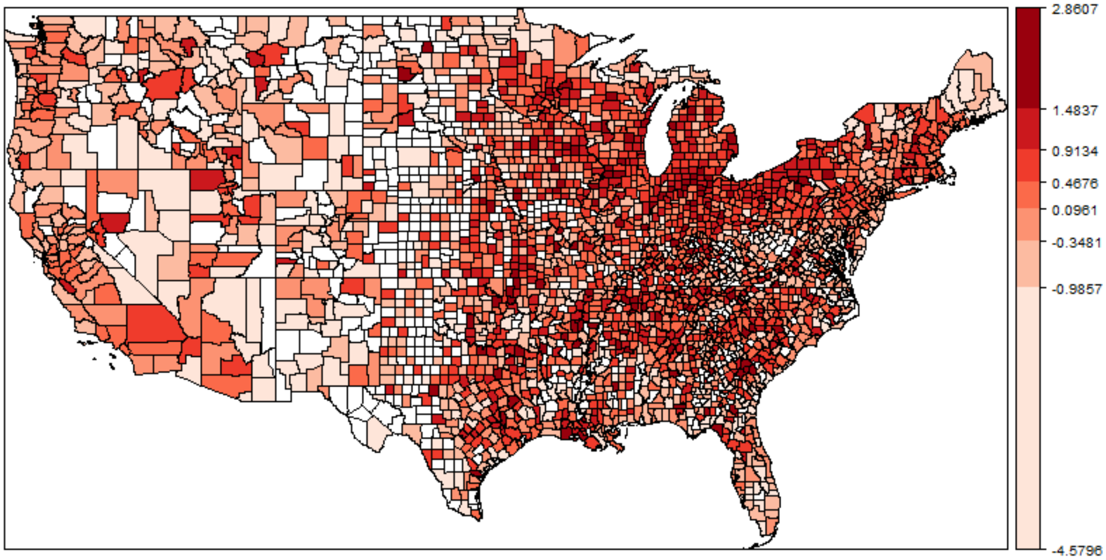

Figure 2.2: The Distribution of Standardized $\log(LQ)$ of Fabricated Metal Manufacturing

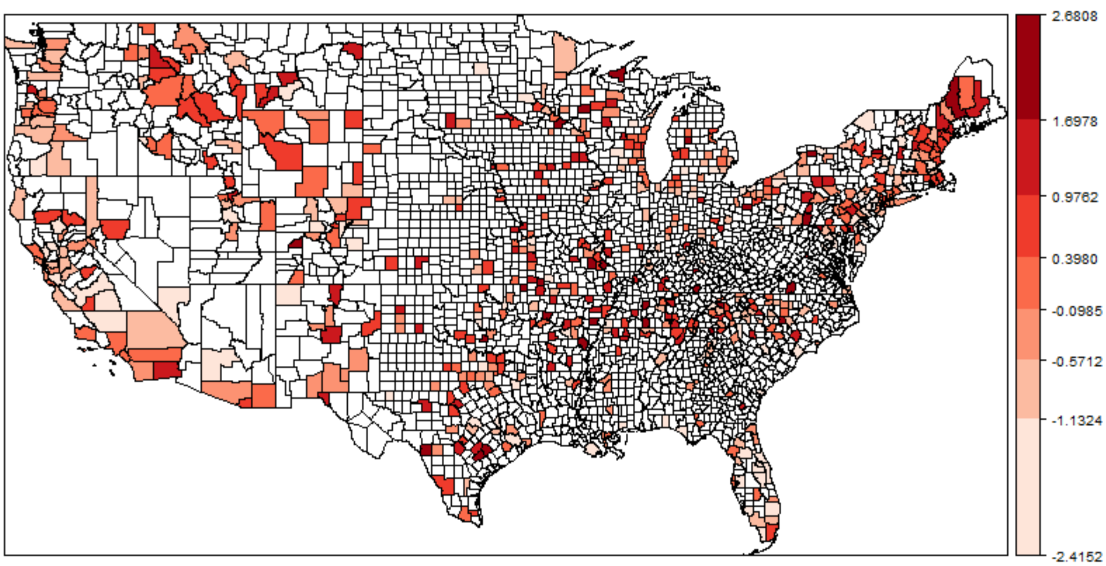

Figure 2.3: The Distribution of Standardized $\log(LQ)$ of Leather & Allied Product Manufacturing

method contributes to the SLQ approach in O'Donoghue and Gleave (2004) by relaxing the normality assumption. An analysis of agglomeration of the 2002 U.S manufacturing industries illustrates the success of this method in identifying agglomeration of an industry for which the LQ is not normally distributed. However, this method does not address other issues concerned with measuring agglomeration. Searching for an agglomeration index satisfying the set of properties initially proposed by Duranton and Overman (2005) is a task for future studies.

The bootstrap method for identifying agglomeration can serve as the starting point for other related studies. The SLQ indices can be readily fed into a spatial econometric regression model in which the spatial weight matrix requires the dependent variable to be region-specific. The spatial econometric model can study the spatial spillover effect of the SLQ from the surrounding regions which may result from estimation errors due to the modifiable areal unit problem. Also, using relatively accurate cut-off values obtained from the bootstrap method, we can divide regions into two contrasting groups, one with agglomeration and another without agglomeration, and then examine the distinctive industrial and local characteristics within each group which can determine the formation of agglomeration.

Chapter 3

Testing the Role of Input-Output Linkages in Industrial Agglomeration with the Spatial Econometric and Input-Output Model

3.1 Introduction

Industrial agglomeration is an economic phenomenon with two dimensions. The first dimension is geographic space. Agglomeration refers to the fact that a large variety of industries locate proximately in space, which gives industries advantages, such as easy communication, knowledge sharing, and reduced transportation costs, which can be considered as spatial spillover effects. The second dimension is industry. The interactions between industries through input-output linkages encourage industries to cluster in a region, which can be considered as inter-industry effects. A study on industrial agglomeration should address both the space and industry dimensions.

This paper uses an integrated spatial econometric and input-output model to meet this end. The integrated spatial econometric and input-output model combines the strengths of spatial econometrics and the input-output model. Spatial econometrics is capable of accounting for spatial spillover effects between regions, while the input-output model is capable of describing the detailed inter-industry relationships. The approach to integrating spatial econometrics and the input-output model is to apply the embedding strategy in the literature of the integrated econometric and input-output model (Rey and Dev, 1997; Rey, 2000) through the spatial intermediate demand variable (SIDV).

The nature of the dataset and the construction of the SIDV suggest building a spatial pseudo-panel Durbin model, estimated by "borrowing" the estimation method of the spatial panel time fixed effects model. A pseudo-panel model means that the model is not a panel data model in the normal sense, which usually has observations on individual spatial units over several time periods. However, the econometric model in this paper preserves the essential form of a panel data model since each spatial unit has observations across several industries instead of time periods. With time fixed effects replaced with industry fixed effects, I name the model as the spatial pseudo-panel industry fixed effects model. The spatial pseudo-panel model is then estimated through the maximum likelihood estimation method used in a spatial panel time fixed effects model. The benefits of applying this trick in estimation are to make full use of information in the dataset, take into account variations along both spatial and industry dimensions, control for industry heterogeneity, and enhance the efficiency of estimation.

Understanding the role of input-output linkages on industrial agglomeration is of practical importance. Regional economic policy makers often select some target industries to attract into their regions, giving these industries tax benefits, facilitating land acquisition, and providing preferential public services. The effectiveness of these policies greatly hinges on the strength of input-output linkages between industries. The settlement of an industry with strong input-output linkages in a region may trigger a self-reinforcing process, leading to more industries agglomerating in that region, as indicated by regional economic theories, such as the growth-pole theory (Perroux, 1950), the cluster theory (Porter, 1998), and New Economic Geography (Krugman and Venables, 1995; Venables, 1996).

3.2 Literature Review

Economic theories have emphasized the role of input-output linkages in the formation of industrial agglomeration for a long time. von Thünen (1826) observes that it is efficient for factories to locate closely. Marshall (1890) advances the concept of input sharing of industries as an ingredient for agglomeration in his famous trinity of Marshallian agglomeration economies (input sharing, labor market pooling, and knowledge spillovers). The vertical linkage models (Krugman and Venables, 1995; Venables, 1996; Robert-Nicoud, 2006) in New Economic Geography (NEG) formalize the analysis of input-output linkages and agglomeration under the general equilibrium framework using the Dixit-Stiglitz monopolistic competition model.

In empirical studies, a straightforward approach to examining the role of input-output linkages on industrial agglomeration is to regress an agglomeration index on a set of explanatory variables that are chosen according to relevant theories. Head and Mayer (2004) summarize this approach as the concentration regressions, which can be considered as the reduced-form approach to testing the NEG theories. The general form of the concentration regressions is as follows,

$$CONC_s = a + bTRCOSTS_s + cIRS_s + dLINKAGES_s + \ldots + e_s \tag{3.1}$$

The dependent variable, $CONC_s$, is a particular agglomeration index for industry s. $TRCOSTS_s$ and IRS_s are proxies for trade costs and the degree of increasing returns, respectively. $LINKAGES_s$ measures the industry's reliance on intermediate inputs. Additional variables can be included to control other factors for agglomeration.

This paper follows the literature of the concentration regressions with an emphasis on the effect of input-output linkages on industrial agglomeration. Table 3.1 lists all the existing empirical studies pertaining to the concentration regression category and having explanatory variables accounting for input-output linkages. Even though the idea of the concentration regression is simple, the number of studies concerning the role of input-output linkages in industrial agglomeration is small. The papers in Table 3.1 are reviewed by answering the following questions. What are the purposes of these studies? What is their estimation method? What are the measures for agglomeration and input-output linkages?

All papers, except Mion (2004), set the goal of testing the sources of industrial agglomeration based on Marshallian agglomeration economies. Rosenthal and Strange (2001), Glaeser and Kerr (2009), and Ellison et al. (2010) attempt to distinguish three sources of Marshallian agglomeration economies by controlling other factors such as natural endowments for U.S. manufacturing industries. Dumais et al. (1997) introduce a dynamic process to examine the effects of Marshallian agglomeration economies on the change of manufacturing employment in U.S. states. Jofre-Monseny et al. (2011) develop a discrete choice model to examine Mashallian agglomeration economies for manufacturing industries in Spain. Overman and Puga (2010) focus on the labor market pooling effect and include input-output linkages as a control variable for manufacturing industries in the U.K. Different from these studies, Mion (2004) directly bases his empirical model on NEG and tests the effects of final demand and input-output linkages on agglomeration.

As for the estimation methods, most of the papers estimate linear regression models using ordinary least square (OLS) estimation. Some studies include fixed effects for industries and/or spatial units (Dumais et al., 1997; Rosenthal and Strange, 2001; Glaeser and Kerr, 2009; Ellison et al., 2010; Overman and Puga, 2010). Mion (2004) estimates a Tobit model with the maximum likelihood (ML) method to account for possible bias caused by the censored data problem in the dependent variable. Jofre-Monseny et al. (2011) employ an innovative method in which a Poisson regression model is derived from the discrete choice model and estimated using the ML method. However, none of the existing studies use spatial

Table 3.1: The Summary of Literature on the Effect of Input-Output Linkages on Agglomeration

Authors, Year	**Main Purpose**	**Estimation Method**	**Measures for Agglomeration**	**Measures for I-O Linkages**
Dumais et al. (1997)	The authors regress the changes in employment share on Marshallian agglomeration forces	linear regressions using OLS with state and industry fixed effects; Non-spatial	$\triangle E^j_{ist}/E_{it}$ and $log(1+\triangle E^j_{ist})$	$input_{ist} \equiv \Sigma_{j\neq i} I_{ji} E_{jst}/E_{jt}$ and $output_{ist} \equiv \Sigma_{j\neq i} O_{ji} E_{jst}/E_{jt}$
Rosenthal and Strange (2001)	Examine the determinants of agglomeration economies for U.S. manufacturing industries, including Marshallian agglomeration economies	linear regressions using OLS with industry fixed effects; Non-spatial	the EG index	energy, natural resources, water, manufactured inputs, non-manufactured inputs as shares of shipment
Mion (2004)	Examines the role of agglomeration externalities stemming from input-output linkages in the location of U.S. manufacturing plants based on NEG theories	the Tobit model using MLE; Spatially lagged independent variables	$Spec_{ci} \equiv ln(1+\frac{emp_{ci}/emp_c}{emp_i/emp})$	$IntD_{ci} = ln(1+\Sigma_k \mu^{c,k} r_k emp_{ki})$, $IntS_{ci} = ln(1+\Sigma_k \mu^{k,c} r_k emp_{ki})$
Glaeser and Kerr (2009)	Examine the impact of local conditions on entrepreneurial rates, including Marshallian agglomeration economies	linear regressions using OLS with city and industry fixed effects; Non-spatial	Log entry employment of new firms for city-industries	$Input_{ci} = -\sum_k \lvert Input_{i\leftarrow k} - \frac{E_{kc}}{E_c}\rvert$, $Output_{ci} = \frac{\left[\sum_k Output_{i\rightarrow k}\frac{E_{kc}}{E_c}\right]}{\left[\sum_k Output_{\cdot\rightarrow k}\frac{E_{kc}}{E_c}\right]}$
Ellison et al. (2010)	Examine the sources of coagglomeration of industries in U.S. manufacturing based on Marshallian agglomeration economies	linear regression using OLS; Non-spatial	EG coagglomeration index and the Duranton and Overmans index	$Input_{ij} = max\{inputs_{i\rightarrow j}, inputs_{j\rightarrow i}\}$, $Output_{ij} = max\{Outputs_{i\rightarrow j}, Outputs_{j\rightarrow i}\}$, $InputOutput_{ij} = max\{inputs_{ij}, Outputs_{ij}\}$
Overman and Puga (2010)	Examine the importance of labor pooling for agglomeration based on Marshallian agglomeration economies	Panel model with industry fixed effects; Non-spatial	the EG index	A set of production factors as share of inputs; IO weighted EG index
Jofre-Monseny et al. (2011)	Examine three Marshallian agglomeration forces in determining location of new firms in Manufacturing in Spain	Poisson regression using MLE; Non-spatial	The number of new firms	$input_{ic} = \sum_{j\neg i}\left(W^I_{ij} L_{cj}\right)$ where $W^I_{ij} = \frac{inputs_{i\rightarrow j}}{totalinputs_i}$ $onput_{ic} = \sum_{j\neg i}\left(W^O_{ij} L_{cj}\right)$ where $W^O_{ij} = \frac{outputs_{i\rightarrow j}}{totaloutputs_i}$

econometric models by including a spatial lag of the dependent variable. The importance of a spatial lag model is explained in Section 3.3.

A prerequisite for the concentration regressions is to find an agglomeration index.Three studies in Table 3.1 use the Ellison-Glaeser (EG) index to measure agglomeration (Rosenthal and Strange, 2001; Ellison et al., 2010; Overman and Puga, 2010). The EG index is derived by Ellison and Glaeser (1997) based on the discrete choice model, designed to correct the bias caused by industry structure on the agglomeration measurement. However, the computation process of the EG index involves the summation over regions, resulting in losing the spatial dimension, which is indispensable in spatial econometric models. Other studies in Table 3.1 compute agglomeration measures that possess both spatial and industry dimensions. Mion (2004) uses the logarithmically transformed location quotient, which is also used the dependent variable in this paper.

Another key variable of interest is the measure of input-output linkages. Rosenthal and Strange (2001) use the shares of energy, natural resources, water, manufactured inputs, and non-manufactured inputs in total shipment to measure the dependency of an industry on natural endowment and estimate their individual effects on agglomeration. Ellison et al. (2010) obtain summarized indices for input-output linages from coefficients in the 1987 benchmark input-output tables of the Bureau of Economic Analysis in the United States. The shortcoming of these measures is that they only have the industry dimension without any locational reference. In contrast, Dumais et al. (1997), Mion (2004), Glaeser and Kerr (2009) and Jofre-Monseny et al. (2011) construct the measures of input-output linkages by combining the input-output tables with regional industry employment, resulting in a measure with both spatial and industry dimensions, which is desirable in this paper.

Reviewing existing empirical research concerning agglomeration and input-output linkages suggests two directions to extend the literature. The first direction is to use spatial econometrics to account for spatial spillover effects. Industrial agglomeration involves spatial distribution of industry production across many regions. Behrens and Thisse (2007) advocate using spatial econometrics to deal with multi-regional problems. However, no papers in Table 3.1 use spatial econometrics except Mion (2004) who includes only spatial lags of independent variables, not the spatial lag of the dependent variable. The inclusion of the spatial lag of dependent variable is critical for modeling the spatial interdependence, and it is also important for accounting for the measurement errors by using a discrete agglomeration index, as the case in all studies in Table 3.1. The measurement errors come from the modifiable area unit problem (MAUP) (Openshaw, 1981), which makes a discrete agglomeration index biased when industries actually agglomerate across the boundaries of spatial units.

The second direction is to integrate spatial econometrics and input-output model. Rey (1998) classifies three integration strategies: embedding, linking, and coupling. Essentially, the methodology of studies in Table 3.1 belongs to the embedding strategy as they all create variables using the input-output matrix, which are then embedded into an econometric model. In the literature on the integrated econometric and input-output model, a variable, called the intermediate demand variable(IDV)(Moghadam and Ballard, 1988; Rey and Jackson, 1999), serves as the building block of the embedding strategy. Further, Rey (2000) advocates extending the integrated econometric and input-output model by including spatial econometrics to model interregional dependence, for which the spatial intermediate demand variable (SIDV)(Rey and Dev, 1997) is a channel to implement the embedding strategy.

3.3 Methodology

As mentioned in Section 3.1, the econometric model is a spatial pseudo-panel industry fixed effects Durbin model. The motivation of estimating such a model arises mainly from three considerations: the nature of data, the way to construct key variables, and, most importantly, the necessity of studying industrial agglomeration with respect to both spatial and industry dimensions.

3.3.1 Data Sources

I obtain industry employment data for U.S. counties from the 2002 County Business Patterns (CBP) imputed by using the method of Isserman and Westervelt (2006). The CBP of the U.S. Bureau of Census provides comprehensive annual records on employment, payroll, and the number of establishments by detailed industry for all counties in the United States. However, the nondisclosure problem of the CBP limits its usefulness. To construct variables used in the model, a complete CBP dataset is indispensable. To deal with the nondisclosure problem, Isserman and Westervelt (2006) propose a method to impute the withheld data, taking advantage of the hierarchical structure of the CBP dataset. I download the complete imputed 2002 CBP dataset[1] and use employment of manufacturing industries with 3-digit NAICS codes[2] at the county level to compute location quotients and the intermediate demand variable.

The data format of the CBP dataset is analogous to a panel data model with two dimensions, counties and industries. With the dummy variables for industries substituting for the dummy variables for time periods, the model can be estimated as if it were a spatial panel time fixed effects model. Moreover, I make the structure of the panel data balanced in that each county has the same number of industries, imposing zeros for employment of industries that do not actually exist in the county. Balancing the panel data makes the spatial weight matrix remain the same for each industry.

The data on industry employment from the CBP dataset is used to compute the location quotient(LQ) and the spatial intermediate demand variable (SIDV). Besides the CBP dataset, I collect county data from other sources, including the Bureau of Economic Analysis (BEA) for the 2002 input-output table at the summary level, USA Counties™ of the Census Bureau for demographic variables, and the Economic Research Service of the U.S. De-

[1]The dataset is downloaded from `www.wholedata.com`. Unfortunately, this website has been shut down recently.

[2]To compute the intermediate demand variable, I use nearly all 3-digit NAICS industries, except for several 2-digit NAICS code industries in BEA's 2002 input-output table at the summary level.

partment of Agriculture (USDA) for the natural amenities scale and the 2003 rural-urban continuum codes. The spatial weight matrix is created using the cartographic boundary files from the Census Bureau.

3.3.2 Variable Construction

The Dependent Variable

The dependent variable in the model is the logarithmically transformed location quotient (LQ). Let $i = 1, 2, \ldots, S$ denote industries and $r = 1, 2, \ldots, N$ denote regions, then the LQ of industry i in region r is defined as

$$LQ_{ir} = \frac{E_{ir}/E_{i*}}{E_{*r}/E_{**}}$$

where E_{ir} is the employment of industry i in region r, E_{i*} is the total employment of industry i in all regions, E_{*r} is the total employment of all industries in region r, and E_{**} is the total employment of the overall economy. I use logarithmic transformation on the LQ to alleviate its problems of influential observations and skewness. For observations for which their LQs are zero, I add a small positive number to prevent the logarithmic transformation from resulting in negative infinity. Thus, the dependent variable is $log(LQ + \epsilon)$, where $\epsilon = \frac{1}{10} min \{LQ : LQ > 0\}$.[3]

Independent Variables

The intermediate demand variable (IDV) is the key variable to implement the embedding strategy of the integrated econometric and input-output model. The IDV is first advanced

[3] Mion (2004) uses $log(LQ + 1)$ as the dependent variable. However, if the LQ is zero for an industry in a county, then $LQ + 1$ will bring up the level of the LQ near to the national average, which is implausible. In fact, I experiment with different values for ϵ by dividing the minimum nonzero LQ by 10 to the power of 1 to 10. The power of 10 gives log(LQ) the smallest standard deviation. Also, the signs of the estimated coefficients in the econometric model are robust for ϵ.

by Moghadam and Ballard (1988) as[4]

$$IDV_{ir} = \sum_{j=1}^{S} a_{ij} E_{jr}, \text{ for } r = 1, \ldots, N. \tag{3.2}$$

The IDV of industry i in region r is the weighted sum of employment of all other industries in region r. The weight a_{ij} represents the share of inputs from industry i in the output of industry j at the national level, i.e. the (i, j) coefficient in the input-output matrix.

Rey and Dev (1997) and Rey (2000) advocate using the spatial intermediate demand variable (SIDV) to introduce spatial linkages into the model. The SIDV is defined as

$$SIDV_{ir} = \sum_{s \neq r} \sum_{j} \phi_{rs} \gamma_{ij}^{s} a_{ij} E_{js}, \text{ for } r = 1, \ldots, N \tag{3.3}$$

where ϕ_{rs} is the freeness of trade between regions r and s, which can be estimated as a decaying function of the distance between r and s, denoted as d_{rs}. In Rey and Dev (1997) $\phi_{rs} = 1/d_{rs}$. γ_{ij}^{s} is the import propensity of industry j in region s for intermediate inputs produced by industry i elsewhere. For simplicity, I assume $\gamma_{ij}^{s} = 1$ for all regions and industries. Rey and Jackson (1999) argue that the original IDV specification needs to be adjusted by labor productivity for each industry. Therefore, in this paper $SIDV_{ir}$ has the following specification,

$$SIDV_{ir} = \sum_{s \neq r} \sum_{j} \frac{1}{d_{rs}} \varphi_{ij} a_{ij} E_{js}, \text{ for } r = 1, \ldots, N \tag{3.4}$$

where $\varphi_{ij} \equiv l_j / l_i$ is the adjustment factor of labor productivity of industry i and j.

Given the form of Equation (3.4), I find that the SIDV can be conveniently computed using an inverse-distance spatial weight matrix. An inverse-distance spatial weight matrix,

[4]The IDV in Moghadam and Ballard (1988), Rey and Dev (1997), and Rey (2000) has a time dimension to introduce dynamics into the input-output model. However, I only have data for one year, thereby I omit the subscript t of the IDV without loss of clarity. Also, the summation should be over $j \neq i$ in a rigorous sense. Since intra-industry trade is possible, especially for industry aggregation into 3-digit NAICS codes in the I-O table, the summation over all industries is reasonable.

denoted as $\boldsymbol{W}$, is an $N \times N$ matrix that has off-diagonal elements $w_{rs} = 1/d_{rs}$ for $r \neq s$ and diagonal elements of zeros, i.e. $w_{rr} = 0$. Define the local IDV for industry i in region r as $\sum_j \varphi_{ij} a_{ij} E_{jr}$. Then, from Equation (3.4), $SIDV_{ir} = \sum_{s \neq r} w_{rs} \sum_j \varphi_{ij} a_{ij} E_{js} = \sum_{s \neq r} w_{rs} IDV_{is}$. In matrix notation, let $\boldsymbol{IDV}_i$ be an $N \times 1$ vector of the IDV for industry i in all regions. Then , $\boldsymbol{SIDV}_i = \boldsymbol{W} \cdot \boldsymbol{IDV}_i$.

Besides input-output linkages, another important factor to explain industrial agglomeration is market potential, which is a variable describing the strength of final demand for industries. I use the nominal market potential (NMP) variable (Harris, 1954; Head and Mayer, 2004) to represent the market potential. The NMP for industry i in region s is defined as $NMP_{ir} = \sum_s \phi_{rs} \mu_{is} Y_s$, where μ_{is} is the share in personal income of consumption for industry i in region s, and Y_s is personal income in region s. Similar to the IDV, the NMP can also be constructed using the inverse-distance spatial weight matrix, $\boldsymbol{W}$, if ϕ_{rs} is approximated by $1/d_{rs}$. Define the local market potential for industry i in region r as $MP_{ir} = \mu_{ir} Y_r$, then $NMP_{ir} = \sum_{s \neq r} w_{rs} \mu_{is} Y_s = \sum_{s \neq r} w_{rs} MP_{is}$. In matrix notation, $\boldsymbol{MP}_i$ is the vector of local market potentials for industry i and $\boldsymbol{W} \cdot \boldsymbol{MP}_i$ is its spatial lag. In this paper, I assume μ_{is} is identical for all regions, i.e., $\mu_{is} = \mu_i$ for all $s = 1, 2, \ldots, N$, and is approximated by the share of industry consumption in total personal income at the national level, for which data are obtained from BEA's 2002 benchmark Use table. And the data for Y_s for each county is collected from USA Counties™ of the Census Bureau.

Additionally, I include two groups of control variables. The first group represents the exogenous natural endowment of regions, consisting of three variables: a natural amenities scale, a dummy variable for coastal counties, and a dummy variable for metro counties. The natural amenities scale is a measure of the physical characteristics of an area. The scale was constructed by combining six measures of climate, topography, and water area that reflect environmental qualities. The definition of coastal counties is provided by the Strategic Environmental Assessments Division of the National Oceanic and Atmospheric Administration. The coastal dummy variable takes the value of one for coastal counties and zero otherwise. The dummy variable for metro counties takes the value of one if the 2003 rural-urban continuum code computed by USDA is less than four, otherwise being zero for nonmetro counties. The second group consists of demographic variables for counties, including the growth rate of population from 1980 to 2000, the percentage of persons who are 25 years and over with a Bachelor's or higher degree in 2000, and the number of violent crimes relative to the national number in 2000. Table 3.2 shows the summary statistics of all variables in the model and the inverse-distance spatial weight matrix.

Table 3.2: Descriptive Statistics of Variables

Variable	Unit	Min	Max	Mean	Median	Std
LQ	national level=1	0.00	259.67	1.38	0.04	5.57
IDV	thousand person	0.00	46.06	0.15	0.02	0.66
MP	million dollors	0.00	8912.58	19.18	1.77	122.88
Amenity	z-scores	-6.40	11.17	0.05	-0.15	2.29
PopGrowth	percentage	-1.86	25.12	0.72	0.40	1.53
Bachelor	percentage	0.05	0.60	0.16	0.14	0.08
Crime	percentage	0.00	6.67	0.03	0.00	0.19
	Number of Nonzero Links	Percentage of Nonzero Links	Min. Number of Neighbors	Max. Number of Neighbors	Ave. Number of Neighbors	
W	247,594	2.65	3.00	169.00	81.05	

(1) Number of observations: 64,155 (3,055 counties in 48 lower states and 21 industries). Independent cities in VA are included in their locating counties. The number of coastal counties is 621. The number of metropolitan counties is 1,052.
(2) Data Sources: The County Business Pattern, USA Counties$^{\mathrm{TM}}$ and the cartographic boundary files from the Census Bureau, the Economic Research Service of the United States Department of Agriculture.
(3) W is the inverse-distance spatial weight matrix, non-standardized, being constructed using the cut-off distance that is 1.5 times as long as the distance ensuring each county to have at least one neighbor.

3.3.3 Model Specification and the Estimation Method

The approach to constructing the SIDV and NMP as spatial lags of IDV and MP dictates the model specification of a spatial Durbin model, estimated as a spatial panel industry fixed-effects model. For each industry $i = 1, 2, \ldots, S$, the model is

$$\boldsymbol{y}_i = \rho \boldsymbol{W} \boldsymbol{y}_i + \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{W} \boldsymbol{X}_i \boldsymbol{\theta} + \boldsymbol{\iota}_N \eta_i + \boldsymbol{\varepsilon}_i, \text{ and } \boldsymbol{\varepsilon}_i \sim IID(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I}_N) \tag{3.5}$$

where $\boldsymbol{y}_i$ is an $N \times 1$ vector of the dependent variable, $\boldsymbol{W}$ is the $N \times N$ inverse-distance spatial weight matrix, ρ is the spatial autoregressive coefficient. $\boldsymbol{X}_i$ is an $N \times K$ matrix of explanatory variables (i.e. IDV, MP, and control variables) excluding the constant term, and $\boldsymbol{\beta}$ is the $K \times 1$ coefficient vector. $\boldsymbol{\iota}_N \eta_i$ is the industry fixed-effects dummy variable, where $\boldsymbol{\iota}_N$ is an $N \times 1$ vector of ones, and $\boldsymbol{\varepsilon}_i$ denotes disturbances assumed to be independently and identically distributed with a mean of zero and variance of σ_ε^2.

Stacking Equation (3.5) over all industries, the model can be expressed in the panel data model format,

$$\boldsymbol{y} = \rho \widetilde{\boldsymbol{W}} \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \widetilde{\boldsymbol{W}} \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{E}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \text{ and } \boldsymbol{\varepsilon} \sim IID(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I}_{NS}) \tag{3.6}$$

where $\widetilde{\boldsymbol{W}}$ is the spatial weight matrix for the panel data, i.e. $\widetilde{\boldsymbol{W}} = \boldsymbol{I}_S \otimes \boldsymbol{W}$, the Kronecker product of the $S \times S$ identity matrix and the spatial weight matrix $\boldsymbol{W}$. $\boldsymbol{y}, \boldsymbol{u},$ and $\boldsymbol{\varepsilon}$ are $NS \times 1$ vectors, $\boldsymbol{X}$ is an $NS \times K$ matrix for explanatory variables, $\boldsymbol{E} = \boldsymbol{I}_S \otimes \boldsymbol{\iota}_N$ is an $NS \times S$ matrix and $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_S)'$ is an $S \times 1$ vector for industry fixed effects. According to Anselin et al. (2008) and Elhorst (2010b), omitting the leading constant, the log-likelihood function of Equation (3.6) is given by

$$\log L = -\frac{NS}{2} \log \sigma_\varepsilon^2 + S \log |\boldsymbol{I}_N - \rho \boldsymbol{W}| - \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^{S} \boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_i \tag{3.7}$$

where $\boldsymbol{\varepsilon}_i = \boldsymbol{y}_i - \rho \boldsymbol{W} \boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{W} \boldsymbol{X}_i \boldsymbol{\theta} - \boldsymbol{\iota}_N \eta_i$. Maximizing Equation (3.7) with respect to coefficients yields the estimation results.

3.4 Estimation

Table 3.3 shows the estimation results of three groups of models. The first two groups of models (Models 1-3 and Models 4-6) are the spatial Durbin model (SDM). The IDV and MP are in their original form in Models 1 to 3, while they are in the logarithmic form in Models 4 to 6. The first models in each group (Models 1 and 4) contains the IDV and MP along with their spatial lags. The second models in each group (Model 2 and 5) add control variables for natural endowments. The third models in each group (Model 3 and 6) add control variables for demographic factors. Model 7 is a spatial autoregressive model (SAR) in which the IDV and W*IDV are combined to form the SIDV, and MP and W*MP are combined to form the NMP. According to the definition of the effects estimates in LeSage and Pace (2010), the results for the direct, indirect, and total effects of explanatory variables in Models 3, 6, and 7 are shown in Table 3.4.

The estimation results confirm the positive role of input-output linkages in agglomeration. The coefficients on the IDV in all models are significantly positive at the 1% levels. The coefficients on the spatial lags of the IDV are significantly positive in Models 1-3, but they are significantly negative in Models 4-6. The logarithmic transformation has an impact on the estimation results. The possible reason for this may be that the sum of log(IDV) and W*log(IDV) is not equal to log(IDV+W*IDV), altering the computation of the SIDV as the sum of the IDV and its spatial lag, W*IDV. The coefficient on the SIDV in model 7 is significantly positive at the 1% level.

Despite the opposite signs of the coefficients on W*IDV and W*log(IDV), the effects estimates of all variants of the IDV are significantly positive in all models. The direct effects of the IDV, log(IDV) and SIDV on log(LQ) are 0.4635 in Model 3, 2.1329 in Model 6, and 0.0642 in Model 7. Take the result of Model 3 and the chemical manufacturing industry (NAICS 325) in Monongalia County, WV as a concrete example. The LQ for chemical manufacturing in the county in 2002 is 4.51, indicating that this industry is concentrated in the county. Consider an increase in the IDV by one unit for the chemical industry in this county, which is a thousand more employees in other industries that have input-output linkages with the chemical industry in the county. With other things being unchanged, the LQ for chemical industry in the county will increase by 46.35% to the level of 6.6, implying greater concentration.

Table 3.3: Estimation Resutls of Cofficients in Various Models

Model Types		SDM			SDM		SAR
Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
IDV	0.5522***	0.4638***	0.4541***	—	—	—	—
	(19.7681)	(16.9643)	(14.2524)				
MP	0.0008***	0.0005***	0.0004***	—	—	—	—
	(5.2079)	(3.4796)	(2.6144)				
log(IDV)	—	—	—	2.1009***	2.1124***	2.1239***	—
				(118.3384)	(115.8085)	(113.7249)	
log(MP)	—	—	—	-1.0172***	-1.1096***	-1.0878***	—
				(-46.7552)	(-45.2433)	(-40.5906)	
SIDV	—	—	—	—	—	—	0.0641***
							(32.7885)
NMP	—	—	—	—	—	—	0.0000
							(0.6215)
Amenity	—	-0.0221**	-0.0576***	—	-0.0264***	-0.0242***	-0.1509***
		(-2.1359)	(-5.2486)		(-2.9649)	(-2.6144)	(-20.1169)
Coastal	—	0.7491***	0.9430***	—	-0.0568	0.0980*	0.1633***
		(12.8836)	(15.7923)		(-1.1335)	(1.9241)	(4.0230)
Metro	—	1.2595***	0.9443***	—	0.3502***	0.3175***	1.3085***
		(35.4853)	(24.3838)		(10.1787)	(9.0751)	(34.6385)
PopGrowth	—	—	0.1712***	—	—	0.1128***	0.2352***
			(13.0356)			(10.0735)	(19.7326)
Bachelor	—	—	3.4316***	—	—	-1.2334***	3.6329***
			(14.6533)			(-5.6985)	(16.0607)
Crime	—	—	-0.0535	—	—	-0.7602***	0.9022***
			(-0.4795)			(-10.5669)	(10.7469)
W*IDV	0.0815***	0.0149***	0.0079***	—	—	—	—
	(38.5018)	(8.2748)	(4.0063)				
W*MP	0.0001***	-0.0000***	-0.0001***	—	—	—	—
	(6.6243)	(-4.7263)	(-5.8186)				
W*log(IDV)	—	—	—	-0.0072***	-0.0075***	-0.0052***	—
				(-63.4906)	(-52.9848)	(-29.1474)	
W*log(MP)	—	—	—	0.0015***	0.0009***	0.0009***	—
				(8.1082)	(4.7008)	(4.5850)	
W*Amenity	—	-0.0021***	-0.0014***	—	0.0005**	0.0010***	—
		(-8.0378)	(-4.1985)		(2.0918)	(3.7011)	
W*Coastal	—	-0.0214***	-0.0249***	—	-0.0036***	-0.0081***	—
		(-14.8642)	(-16.8583)		(-2.8968)	(-6.4292)	
W*Metro	—	0.0621***	0.0437***	—	0.0072***	-0.0090***	—
		(40.5806)	(16.9670)		(5.6046)	(-4.0619)	
W*PopGrowth	—	—	-0.0006	—	—	-0.0016***	—
			(-1.1854)			(-3.4151)	
W*Bachelor	—	—	0.1268***	—	—	0.1073***	—
			(15.2923)			(12.0826)	
W*Crime	—	—	-0.0347***	—	—	-0.0054	—
			(-3.0116)			(-0.5957)	
ρ	0.0030***	0.0060***	0.0070***	0.0070***	0.0080***	0.0070***	0.0030***
	(14.2405)	(92.9795)	(211.3044)	(302.9170)	(2050.9156)	(260.0239)	(14.8933)
R^2	0.2292	0.2886	0.2981	0.4986	0.5007	0.5032	0.2662
loglikelihood	-178872	-178245	-178436	-169796	-171118	-169720	-177377

Notes: (1) Significant levels: * for 10%, ** for 5%,and *** for 1%.
(2) The t-statistics is enclosed in the parenthesis.

Table 3.4: Estimation of Direct, Indirect and Total Effects

Effects		Direct Effects	Indirect Effects	Total Effects
Model 3	IDV	0.4635***	1.9831***	2.4467***
		(15.0986)	(5.7668)	(7.1186)
	MP	0.0004**	-0.0088***	-0.0085***
		(2.2975)	(-5.7209)	(-5.5431)
	Amenity	-0.0596***	-0.3256***	-0.3852***
		(-5.4851)	(-6.0817)	(-7.7023)
	Coastal	0.9253***	-3.4393***	-2.5139***
		(15.2821)	(-15.0567)	(-12.2614)
	Metro	0.9952***	9.1516***	10.1468***
		(26.2629)	(20.9860)	(23.3425)
	PopGrowth	0.1718***	0.0899	0.2617***
		(13.3651)	(0.9494)	(2.8456)
	Bachelor	3.5808***	27.2772***	30.8580***
		(15.1656)	(18.5548)	(20.5218)
	Crime	-0.0828	-6.2483***	-6.3312***
		(-0.7339)	(-2.9014)	(-2.9184)
Model 6	log(IDV)	2.1328***	1.5730***	3.7059***
		(115.9015)	(36.3342)	(69.7582)
	log(MP)	-1.0946***	-1.1225***	-2.2171***
		(-42.3041)	(-25.6834)	(-35.7652)
	Amenity	-0.0238***	0.1638***	0.1400***
		(-2.5915)	(3.6216)	(3.3229)
	Coastal	0.0896*	-1.3580***	-1.2684***
		(1.7784)	(-6.8581)	(-7.0068)
	Metro	0.3106***	-1.2786***	-0.9680**
		(8.7443)	(-3.0858)	(-2.3322)
	PopGrowth	0.1118***	-0.1526**	-0.0409
		(10.2892)	(-1.9966)	(-0.5505)
	Bachelor	-1.1210***	18.1876***	17.0666***
		(-5.0532)	(11.5764)	(10.8212)
	Crime	-0.7700***	-1.9377	-2.7077*
		(-10.5527)	(-1.1965)	(-1.6611)
Model 7	SIDV	0.0642***	0.0173***	0.0815***
		(31.9756)	(12.5512)	(41.1598)
	NMP	0.0000	0.0000	0.0000
		(0.6467)	(0.6376)	(0.6454)
	Amenity	-0.1511***	-0.0408***	-0.1918***
		(-21.0786)	(-9.7682)	(-20.1433)
	Coastal	0.1636***	0.0441***	0.2077***
		(4.0289)	(3.7938)	(4.0258)
	Metro	1.3110***	0.3539***	1.6650***
		(34.2326)	(10.0018)	(28.0501)
	PopGrowth	0.2355***	0.0635***	0.2990***
		(19.6493)	(9.9419)	(19.4156)
	Bachelor	3.6252***	0.9785***	4.6037***
		(15.8975)	(8.8061)	(15.2115)
	Crime	0.9058***	0.2445***	1.1503***
		(10.7242)	(7.5153)	(10.5048)

Notes: (1) Significant levels: * for 10%, ** for 5%,and *** for 1%.
(2) The t-statistics are enclosed in the parenthesis.

As for the indirect effects estimates, Model 3 has a stronger indirect effect than the direct effect, while Model 6 and 7 have weaker indirect effects than direct effects. Given the construction of the inverse-distance spatial weight matrix used in estimation, a stronger indirect effects may be reasonable. For example, Monongalia County, WV, has 114 neighbors in the weight matrix, which could imply a large amount of intermediate demand from its neighbors and its neighbor's neighbors. For Model 6, a weaker indirect effect may still be due to the logarithmic transformation. But for Model 7, since the SIDV takes into account intermediate demand from neighboring counties, the direct effect may have already explained most parts of the impacts of the SIDV on log(LQ), leading to a weaker indirect effect estimate. Despite a relatively weak indirect effect in Model 6, an increase of 1%in the IDV from neighboring counties can contribute to an increase of 1.57% in the LQ for an industry in a county. Overall, the total effects, which combine the direct and indirect effects, are significantly positive in all models.

The estimation results for market potential are counterintuitive. While market potential is an important force for agglomeration in theory, as far as manufacturing industries in U.S. counties are concerned, the estimation results does not lend strong evidence for that. In Model 3, the direct effect of the MP is 0.0004, which means that the LQ for an industry can increase by 0.04% if the local final demand for the industry increases by $1 million. Continuing with the example of Monongalia County, the $1 million increase in final demand in the county only results in the chemical industry employment increasing by less than one person. However, the small gain will be overwhelmed by the indirect effect. An increase in market potential in neighboring counties will pull some manufacturing production out of the county. In Model 7, which combines the MP and its spatial lag as the NMP, the direct, indirect, and total effects are all insignificant, further illustrating the counteraction of the local and neighboring final demand markets. Mion (2004) also gets negative coefficients on market potential. He explains that manufacturing activities are pushed aside from the center by sectors characterized by higher transportation costs. Facing higher wages and rents in the center, manufacturing industries, which do not depend on face-to-face interactions, tend to locate closer to intermediate demand markets than to final demand markets.

Robustness Check with Spatial Weight Matrices

All model estimations in Table 3.3 use the inverse-distance spatial weight matrix without any standardization. To test the robustness of the estimation results with respect to the construct of the spatial weight matrix, Model 1 is re-estimated with different weight matrices, as shown in Table 3.5. Using three different weight matrices, the estimated coefficients on the IDV and MP are very close to those in Table 3.3. However, the spatial weight matrix with standardization raises the magnitude of the estimated spatial autoregressive coefficient, implying stronger indirect and total effects estimates. As shown in the lower panel in Table 3.5, with the globally standardized weight matrix (W3), the indirect effect of the IDV is the largest. Nonetheless, using different spatial weight matrices does not change the sign of the effects of the IDV on industrial agglomeration, which proves the robustness of the main findings.

Table 3.5: Estimation with Different Spatial Weight Matrices

		W1	W2	W3
Coefficients	IDV	0.6015	0.5457	0.5796
	MP	0.0007	0.0008	0.0006
	W*IDV	0.1272	5.4845	0.6017
	W*MP	0.0003	0.0037	0.0012
	ρ	0.7890	0.2210	0.4090
Effects Estimates				
Direct Effects	IDV	0.6297	0.5725	0.6520
	MP	0.0008	0.0008	0.0007
Indirect Effects	IDV	2.8341	7.1623	1.3451
	MP	0.0041	0.0049	0.0023
Total Effects	IDV	3.4639	7.7347	1.9971
	MP	0.0049	0.0057	0.0030

Notes: All estimations in this table are based on Model 1.
W1 is the inverse distance spatial weight matrix with row standardization,
W2 is the inverse distance spatial weight matrix with global standardization,
W3 is the five-nearest neighbor spatial weight matrix with row standardization.

Issues with the Estimation Method

While the estimation results validate the role of input-output linkages on agglomeration, issues with the estimation method deserve some caution. The main concern is the endogeneity problem. The LQ and IDV are all computed using data on industry employment, which could inherently entail the endogeneity problem in the estimation. Also, the IDV and MP may be endogenous, determined by other exogenous variables that can also explain industrial agglomeration. Further, the causal relationship between agglomeration and these variables may be reverse. The consequence of the endogeneity problem is that the ML estimates may be inconsistent because the true data generating process is not fully specified merely through the likelihood function (Equation 3.7). A solution to this problem is to use the IV/GMM estimation method, which is proposed by Kelejian and Prucha (1998, 1999) to deal with the existence of ρWy and simplify the computation process in the ML estimation. Mutl and Pfaffermayr (2011) and Millo and Piras (2012) explain the IV/GMM method for spatial panel models. The advantage of the IV/GMM method is that the instrumental variables can be used to handle the endogeneity problem of both the spatially lagged dependent variable and suspicious explanatory variables. The disadvantage of the IV/GMM method is that the estimate of the spatial autoregressive coefficient, ρ, cannot be guaranteed to be within its parameter space, which is $(1/\omega_{min}, 1/\omega_{max})$ where ω_{min} and ω_{max} are the minimum and maximum of the eigenvalues of the spatial weight matrix (Elhorst, 2010b). Given that the estimation results are robust regarding the sign of coefficients on the IDV, the qualitative relationship between input-output linkages and industrial agglomeration is well established, not affected by the endogeneity issue.

3.5 Conclusion

This paper examines the role of input-output linkages between industries in the formation of industrial agglomeration. The spatial and industry dimensions of industrial agglomeration suggest building the model under the framework of the integrated spatial econometric and input-output model through the spatial intermediate demand variable. Taking advantage of the data format and the construction of the spatial intermediate demand variable and the nominal market potential variable, this paper estimates the model as the spatial pseudo-panel industry fixed effects Durbin model. Using manufacturing industries in U.S. counties, the role of input-output linkages is supported. However, the potential endogeneity problem needs to be addressed.

Chapter 4

Exploring the Space-Industry Filter in Regional Industry Models

4.1 Introduction

The modeling techniques of regional economics have been greatly enhanced thanks to the advancement of spatial econometrics. A spatial econometric model with a spatial lagged dependent variable is considered as a formal specification of the equilibrium outcome of a spatial or social interaction process (Anselin et al., 2008). The presence of a spatially lagged dependent variable, $\boldsymbol{Wy}$, where $\boldsymbol{W}$ is a spatial weight matrix, introduces non-linearity stemming from the inter-dependence of the dependent variable. However, such inter-dependence is by no means confined only in the spatial sense. Economic activities are connected in a certain way for which we can construct a weight matrix, say $\boldsymbol{A}$, to depict their inter-dependence. This paper focuses on the inter-dependence of regional industry activity by setting up a spatial econometric model through the space-industry filter.

The space-industry filter can be considered as a composite weight matrix, composed of a spatial weight matrix and an input-output matrix. It originates from the space-time filter introduced by Parent and LeSage (2012) and Debarsy et al. (2012). In those papers, the space-time filter is used to simplify the representation of a dynamic spatial panel data model. The space-time filter can be written as $\boldsymbol{C} \otimes \boldsymbol{B}$, where $\boldsymbol{B} = \boldsymbol{I}_N - \rho \boldsymbol{W}$, $\boldsymbol{C} = I_T - \phi \boldsymbol{L}$, $\boldsymbol{W}$ is a spatial weight matrix, and $\boldsymbol{L}$ is a matrix representing the time-lag operator. ρ and ϕ are the coefficients for spatial and temporal autoregressive processes, respectively. Parent and LeSage (2012) explore a Bayesian MCMC estimation for the dynamic spatial panel data model, distinguishing two specifications depending on whether the first period observation is exogenous or endogenous. Debarsy et al. (2012) introduce a way to interpret the coefficient estimates in the model with a space-time filter. Based on these two papers, the space-industry filter replaces the matrix of the time-lag operator with an input-output matrix that describes the input-output relationship between industries. Composed of a spatial weight matrix and an input-output matrix, the space-industry filter is capable of modeling the general equilibrium in regional industry activity as the outcome of both inter-regional and inter-industry interaction.

I regard the new modeling technique in this paper as a contribution to the literature of the integrated econometric and input-output models. Rey (1998, 2000) reviews works integrating econometrics and the input-output model. Among the three integration strategies, coupling, linking, and embedding, my approach extends the embedding strategy. Unlike previous works, I do not simply embed an explanatory variable created with an input-output matrix, for example, the intermediate demand variable (IDV) (Moghadam and Ballard, 1988; Rey and Dev, 1997; Rey and Jackson, 1999). Instead I embed the input-output matrix directly into the model through the space-industry filter, which is imposed on the dependent variable. The model with a space-industry filter is a Cliff-Ord type spatial econometric model, describing regional and industry inter-dependence in one model setting.

4.2 The Model Specification and Estimation

This section shows how to construct a spatial econometric model with industry fixed effects using the space-industry filter. Regional fixed effects are captured by regional-specific variables in the model instead of using dummy variables. Then I derive a Bayesian MCMC method to estimate the model.

4.2.1 The Model Specification

The dependent variable $\boldsymbol{y}$ is an $NS \times 1$ vector, with y_{ij} representing the activity of industry j in region i. $\boldsymbol{X}$ is an $NS \times K$ matrix of the independent variables with the $K \times 1$ coefficient vector $\boldsymbol{\beta}$. $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_S)'$ is a $S \times 1$ vector for the industry specific effects, which are assumed to be normally distributed, i.e., $\boldsymbol{\eta} \sim N(\mathbf{0}, \sigma_\eta^2 \boldsymbol{I}_S)$. The disturbance $\boldsymbol{\varepsilon}$ is also assumed to be normally distributed, i.e., $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I}_{NS})$, and independent with $\boldsymbol{\eta}$.

Let $\boldsymbol{W}$ be an $N \times N$ spatial weight matrix, and $\boldsymbol{A}$ be an $S \times S$ input-output matrix. Further, define $\boldsymbol{B} = \boldsymbol{I_N} - \rho \boldsymbol{W}$ and , $\boldsymbol{C} = \boldsymbol{I_S} - \phi \boldsymbol{A}$. Then, the space-industry filter is $\boldsymbol{C} \otimes \boldsymbol{B}$, which can be expanded as,

$$\boldsymbol{C} \otimes \boldsymbol{B} = \boldsymbol{I}_{NS} - \rho \boldsymbol{I}_S \otimes \boldsymbol{W} - \phi \boldsymbol{A} \otimes \boldsymbol{I}_N + \rho\phi \boldsymbol{A} \otimes \boldsymbol{W} \tag{4.1}$$

Multiplying the space-industry filter with $\boldsymbol{y}$, the model can be written and re-arranged as follows,

$$\begin{aligned}
(\boldsymbol{C} \otimes \boldsymbol{B})\boldsymbol{y} &= \boldsymbol{X\beta} + \boldsymbol{E\eta} + \boldsymbol{\varepsilon} \\
\boldsymbol{y} &= \rho(\boldsymbol{I}_S \otimes \boldsymbol{W})\boldsymbol{y} + \phi(\boldsymbol{A} \otimes \boldsymbol{I}_N)\boldsymbol{y} - \rho\phi(\boldsymbol{A} \otimes \boldsymbol{W})\boldsymbol{y} + \boldsymbol{X\beta} + \boldsymbol{E\eta} + \boldsymbol{\varepsilon} \\
\boldsymbol{E} &= \boldsymbol{I}_S \otimes \boldsymbol{\iota}_N
\end{aligned} \tag{4.2}$$

where $\rho(\boldsymbol{I}_S \otimes \boldsymbol{W})\boldsymbol{y}$ represents spatial dependence of industry activity in neighboring regions. $\phi(\boldsymbol{A} \otimes \boldsymbol{I}_N)\boldsymbol{y}$ captures inter-industry dependence among all industries in the same region. $\rho\phi(\boldsymbol{A} \otimes \boldsymbol{W})\boldsymbol{y}$ represents the cross-effects combining both regional and inter-industry dependence.

4.2.2 Bayesian Estimation

Following the method in Parent and LeSage (2012), equation (4.2) is estimated by the Bayesian Markov Chain Monte Carlo (MCMC) estimation method. The Bayesian MCMC estimation method relies on the construction of the posterior distribution, which is proportional to the product of the likelihood function and the prior distribution of the parameters. A general form of the posterior distribution is as follows,

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) \times \pi(\boldsymbol{\theta}) \tag{4.3}$$

where $f(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function, $\pi(\boldsymbol{\theta})$ is the joint prior distribution of the parameters, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\eta}', h_\varepsilon, h_\eta, \rho, \phi)'$, where $h_\varepsilon = \sigma_\varepsilon^{-2}$ and $h_\eta = \sigma_\eta^{-2}$, i.e., the precision parameters for $\boldsymbol{\varepsilon}$ and $\boldsymbol{\eta}$, respectively. Given that $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, h_\varepsilon^{-1}\boldsymbol{I}_{NS})$, the likelihood function of equation (4.2) is

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = (2\pi)^{-\frac{NS}{2}} h_\varepsilon^{\frac{NS}{2}} |\boldsymbol{P}| \exp(-\frac{h_\varepsilon}{2}\boldsymbol{e}'\boldsymbol{e}) \tag{4.4}$$

where $\boldsymbol{e} = \boldsymbol{Py} - \boldsymbol{X\beta} - \boldsymbol{E\eta}$, and $\boldsymbol{P} = \boldsymbol{C} \otimes \boldsymbol{B}$. By the property of the Kronecker product, the Jacobian term can be written as, $|\boldsymbol{P}| = |\boldsymbol{C}|^N |\boldsymbol{B}|^S$, which simplifies the computation of $|\boldsymbol{P}|$.

I use the non-informative prior distributions as in Parent and LeSage (2012). The prior distributions for all parameters in the model are given by

$$\begin{aligned}
&\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \boldsymbol{M}_\beta^{-1}),\ \boldsymbol{\eta} \sim N(\boldsymbol{0}, h_\eta^{-1}\boldsymbol{I}_S) \\
&h_\varepsilon \sim G(\frac{\upsilon_0}{2}, \frac{\delta_0}{2}),\ h_\eta \sim G(\frac{\upsilon_1}{2}, \frac{\delta_1}{2}) \\
&\rho \sim U(-1, 1),\ \phi \sim U(-1, 1)
\end{aligned} \tag{4.5}$$

where $G(\cdot)$ is the Gamma distribution and $U(\cdot)$ is the uniform distribution. For the hyperparameters in the prior distributions, $\boldsymbol{\beta}_0, \boldsymbol{M}_\beta, \upsilon_0, \delta_0, \upsilon_1, \delta_1$, I set the prior means of $\boldsymbol{\beta}_0$ to be zero, the variance $\boldsymbol{M}_\beta^{-1}$ to be $10^{12} I_K$, and the parameters for the Gamma priors to be 0.001.

From equations (4.3), (4.4) and (4.5), the posterior distribution for equation (4.2) can be expressed as

$$
\begin{aligned}
f(\boldsymbol{\beta}, \boldsymbol{\eta}, h_\varepsilon, h_\eta, \rho, \phi|\boldsymbol{y}) \propto\ & h_\varepsilon^{\frac{NS}{2}} |\boldsymbol{P}| \exp\left(-\frac{h_\varepsilon}{2}\boldsymbol{e}'\boldsymbol{e}\right) && (4.6)\\
& \times \exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{M}_\beta(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right) && (4.7)\\
& \times h_\varepsilon^{\frac{\upsilon_0}{2}-1} \exp\left(-\frac{\delta_0 h_\varepsilon}{2}\right) && (4.8)\\
& \times h_\eta^{\frac{S}{2}} \exp(-\frac{h_\eta}{2}\boldsymbol{\eta}'\boldsymbol{\eta}) && (4.9)\\
& \times h_\eta^{\frac{\upsilon_1}{2}-1} \exp\left(-\frac{\delta_1 h_\eta}{2}\right) && (4.10)\\
& \times \frac{1}{2} \times \frac{1}{2} && (4.11)
\end{aligned}
$$

In the posterior distribution, equation (4.6) is the likelihood function, ignoring the leading constant. Equations (4.7) and (4.9) are the prior normal distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$, also ignoring the parts that do not include the parameters of interest. Equations (4.8) and (4.10) are the prior Gamma distribution of h_ε and h_η. In the last line, $\frac{1}{2}$ represents the prior uniform distribution of ρ and ϕ.

The Bayesian MCMC method simulates the posterior distribution by generating random samples from the conditional posterior distribution of parameters. I use the Gibbs algorithm to generate samples of $(\boldsymbol{\beta}, \boldsymbol{\eta}, h_\varepsilon, h_\eta)$ and the Metropolis-Hasting algorithm to generate ρ and ϕ. The detailed derivation of the conditional posterior distributions and the algorithm of the Bayesian MCMC estimation can be seen in the appendix.

4.3 Interpretation of Coefficient Estimates

The space-industry filter makes interpretation of coefficient estimates more complicated than that in a traditional spatial econometric model. LeSage and Pace (2010) introduce an approach to interpreting coefficient estimates of explanatory variables in a spatial lag model. They use summary measures of direct, indirect, and total effects to explain the effects of an explanatory variable on the dependent variable within a region, across different regions, and over all regions. However, the space-industry filter includes new "spillover effects" between industries. Thus, it is necessary to devise new summary measures to take into account both the regional and industry relationships. Further, the effects estimates can be partitioned along both spatial and industry dimensions to examine the rate of decay of impacts over space and input-output chains among industries.

4.3.1 Defining Summary Measures for Effects Estimates

The approach to interpreting the coefficient estimates in equation (4.2) follows the idea of LeSage and Pace (2010) by first looking at the form of the partial derivative of the dependent variable with respect to an explanatory variable. Letting $\boldsymbol{Q} = \boldsymbol{P}^{-1}$, equation (4.2) can be re-written as

$$\boldsymbol{y} = \boldsymbol{Q}\left(\boldsymbol{X\beta} + \boldsymbol{E\eta} + \boldsymbol{\varepsilon}\right) \tag{4.12}$$

Then consider the rth explanatory variables, i.e., the rth column in $\boldsymbol{X}$, denoted as $\boldsymbol{x}^r$. Decompose $\boldsymbol{y}$, $\boldsymbol{Q}$, and $\boldsymbol{x}^r$ and re-write equation (4.12) as

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_S \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} & \cdots & \boldsymbol{Q}_{1S} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} & \cdots & \boldsymbol{Q}_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{Q}_{S1} & \boldsymbol{Q}_{S2} & \cdots & \boldsymbol{Q}_{SS} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1^r \\ \boldsymbol{x}_2^r \\ \vdots \\ \boldsymbol{x}_S^r \end{pmatrix} \beta_r + \text{remaining terms} \tag{4.13}$$

where $\boldsymbol{y}_i$ is an $N \times 1$ vector for industry i, $\boldsymbol{Q}_{ij}$ is the (i, j) block of $\boldsymbol{Q}$, which is an $N \times N$ matrix, and $\boldsymbol{x}_i^r$ is an $N \times 1$ vector of the rth explanatory variables for industry i. The partial derivative of $\boldsymbol{y}_i$ with respect to $\boldsymbol{x}_i^r$ is

$$\frac{\partial \boldsymbol{y}_i}{\partial \boldsymbol{x}_i^r} = \boldsymbol{Q}_{ii} \beta_r$$

It means that the diagonal blocks, $\boldsymbol{Q}_{ii}$, can account for the effects of the rth explanatory variable on the dependent variable within the same industry. Looking further inside of $\boldsymbol{Q}_{ii}$, the diagonal elements correspond to the effects within the same industry and the same region, and the off-diagonal elements correspond to the effects within the same industry but across other regions. The partial derivative of $\boldsymbol{y}_i$ with respect to $\boldsymbol{x}_j^r$ is

$$\frac{\partial \boldsymbol{y}_i}{\partial \boldsymbol{x}_j^r} = \boldsymbol{Q}_{ij} \beta_r$$

which means that the off-diagonal blocks, $\boldsymbol{Q}_{ij}$, can account for the effects across different industries and all regions.

Following the idea of the direct, indirect, and total effects in LeSage and Pace (2010), I define similar summary measures shown in Table 4.1. First, define the total space-industry effects as the effects of $\boldsymbol{x}^r$ on $\boldsymbol{y}$ spreading across all regions and all industries, which can be computed as β_r times the sum of all elements in $\boldsymbol{Q}$ divided by NS. Second, define the direct space-industry effects as the effects of $\boldsymbol{x}^r$ on $\boldsymbol{y}$ within the same region and the same industry. The direct space-industry effects can be computed as β_r times the sum of the trace of $\boldsymbol{Q}$ divided by NS. Third, define the direct within-industry effects as the effects of $\boldsymbol{x}^r$ on $\boldsymbol{y}$ within the same industry but across other regions. The direct within-industry effects can be computed as β_r times the sum of all elements in all diagonal blocks divided by NS and then minus the direct space-industry effects. Finally, define the indirect space-time effects as the effects of $\boldsymbol{x}^r$ on $\boldsymbol{y}$ spreading across all other regions and all other industries. The indirect space-industry effects can be computed as the total effects minus the direct space-industry effects minus the direct within-industry effects.

Table 4.1: Definition of the Effects Estimates in the Model with a Space-Industry Filter

Effects names	Formula	Meaning
the total space-industry effects	$\beta_r \boldsymbol{\iota}'_{NS}\boldsymbol{Q}\boldsymbol{\iota}_{NS}/NS$	the effects of $\boldsymbol{x}^r$ on $\boldsymbol{y}$ spreading out across all regions and all industries
the direct space-industry effects	$\beta_r \mathrm{tr}(\boldsymbol{Q})/NS$	the effects of $\boldsymbol{x}^r$ on $\boldsymbol{y}$ within the same region and the same industry
the direct within-industry effects	$\beta_r(\sum_{i=1}^{S}\boldsymbol{\iota}'_N\boldsymbol{Q}_{ii}\boldsymbol{\iota}_N - \mathrm{tr}(\boldsymbol{Q}))/NS$	the effects of $\boldsymbol{x}^r$ on $\boldsymbol{y}$ within an industry but spreading out all other regions
the indirect space-industry effects	$\beta_r \sum_{i=1}^{S}\sum_{j\neq i}^{S}\boldsymbol{\iota}'_N\boldsymbol{Q}_{ij}\boldsymbol{\iota}_N/NS$	the effects of $\boldsymbol{x}^r$ on $\boldsymbol{y}$ spreading out across all other regions and all other industries

As a special case for the row-standardized weight matrices, $\boldsymbol{A}$ and $\boldsymbol{W}$, the total space-industry effects, $\beta_r\boldsymbol{\iota}'_{NS}\boldsymbol{Q}\boldsymbol{\iota}_{NS}/NS$, can take a simple form, using the property that $\boldsymbol{A}\boldsymbol{\iota}_S = \boldsymbol{\iota}_S$ and $\boldsymbol{W}\boldsymbol{\iota}_N = \boldsymbol{\iota}_N$. In this case, the total effects is,

$$\begin{aligned}
(NS)^{-1}\boldsymbol{\iota}'_{NS}\boldsymbol{Q}\boldsymbol{\iota}_{NS}\beta_r &= (NS)^{-1}\boldsymbol{\iota}'_{NS}(\boldsymbol{C}\otimes\boldsymbol{B})^{-1}\boldsymbol{\iota}_{NS}\beta_r \\
&= (NS)^{-1}(\boldsymbol{\iota}_S\otimes\boldsymbol{\iota}_N)'(\boldsymbol{C}^{-1}\otimes\boldsymbol{B}^{-1})(\boldsymbol{\iota}_S\otimes\boldsymbol{\iota}_N)\beta_r \\
&= S^{-1}\left(\boldsymbol{\iota}'_S(\boldsymbol{I_S}-\phi\boldsymbol{A})^{-1})\boldsymbol{\iota}_S\right)\otimes N^{-1}\left(\boldsymbol{\iota}'_N(\boldsymbol{I_N}-\rho\boldsymbol{W})^{-1})\boldsymbol{\iota}_N\right)\beta_r \\
&= S^{-1}\left(\boldsymbol{\iota}'_S(\boldsymbol{I}_S+\phi\boldsymbol{A}+\phi^2\boldsymbol{A}^2+\phi^3\boldsymbol{A}^3+\ldots)\boldsymbol{\iota}_S\right)\otimes \\
&\quad N^{-1}\left(\boldsymbol{\iota}'_N(\boldsymbol{I}_N+\rho\boldsymbol{W}+\rho^2\boldsymbol{W}^2+\rho^3\boldsymbol{W}^3+\ldots)\boldsymbol{\iota}_N\right)\beta_r \\
&= (1-\phi)^{-1}(1-\rho)^{-1}\beta_r
\end{aligned} \tag{4.14}$$

From equation (4.14), it is clear that the total effects of $\boldsymbol{x}^r$ on $\boldsymbol{y}$ incorporate the forces coming from both inter-industry and inter-regional relationships. $(1-\phi)^{-1}$ is the multiplier effects arising from the input-output linkages, and $(1-\rho)^{-1}$ is the spatial spillover effects arising from the spatial linkages.

4.3.2 Spatially-Oriented and IO-Oriented Partitioned Effects

Following the way of partitioning effects estimates by order of spatial neighbors in LeSage and Pace (2010), with the space-industry filter, I partition effects estimates by both order of neighbors and rounds of input-output processes. As spatial spillover effects spread from low-order neighbors to high-order ones, its magnitude is expected to decrease. Similarly, as production goes through several rounds of input-output processes, the magnitude of the multiplier effects is also expected to decrease. The speed of decay by order of neighbors and rounds of input-output processes depends on the structure of the spatial weight matrix and the input-output matrix as well as ρ and ϕ. Examining and comparing the speed between the two kinds of partitioning methods offers more detailed insights about the strength of spatial and input-output linkages than the summary measures defined above.

Spatially-Oriented Partitioned Effects

With the matrix, $\boldsymbol{Q}$, expanded and its expression re-arranged, spatially-oriented partitioned effects can be derived from the following equations,

$$
\begin{aligned}
\boldsymbol{Q} &= [(\boldsymbol{I}_S - \phi \boldsymbol{A}) \otimes (\boldsymbol{I}_N - \rho \boldsymbol{W})]^{-1} \\
&= \boldsymbol{I}_{NS} + \boldsymbol{I}_S \otimes (\rho \boldsymbol{W} + \rho^2 \boldsymbol{W}^2 + \rho^3 \boldsymbol{W}^3 + \cdots) && (4.15) \\
&+ (\phi \boldsymbol{A} + \phi^2 \boldsymbol{A}^2 + \phi^3 \boldsymbol{A}^3 + \cdots) \otimes \boldsymbol{I}_N && (4.16) \\
&+ (\phi \boldsymbol{A} + \phi^2 \boldsymbol{A}^2 + \phi^3 \boldsymbol{A}^3 + \cdots) \otimes \rho \boldsymbol{W} && (4.17) \\
&+ (\phi \boldsymbol{A} + \phi^2 \boldsymbol{A}^2 + \phi^3 \boldsymbol{A}^3 + \cdots) \otimes \rho^2 \boldsymbol{W}^2 && (4.18) \\
&+ (\phi \boldsymbol{A} + \phi^2 \boldsymbol{A}^2 + \phi^3 \boldsymbol{A}^3 + \cdots) \otimes \rho^3 \boldsymbol{W}^3 && (4.19) \\
&+ \cdots\cdots
\end{aligned}
$$

Each line in the equation represents a different meaning, for which I define the direct, indirect and total effects.

Equation (4.15) represents the pure spatial spillover effects as in a common spatial econometric model, without input-output linkages involved. Thus, we get the direct effects of an explanatory variable, X_r, as the sum of the traces of matrices in equation (4.15), multiplied by β_r and divided by the product of the number of regions and the number of industries, NS. The total effects are the sum of all elements of the matrices in equation (4.15), multiplied by β_r and divided by NS. And the indirect effects are the difference of the total effects and direct effects.

Equation (4.16) represents the pure input-output linkages, accounting for the multiplier effects from the first round input-output process and the rest of rounds in the own region. Analogous to the spatial case, the direct effects are the sum of the traces of matrices in equation (4.16), multiplied by β_r and divided by NS, representing the input-output linkages within the own industries. The total effects are the sum of all elements of the matrices in equation (4.16), multiplied by β_r and divided by NS, including all rounds of input-output effects. And the indirect effects are the difference of the total effects and direct effects.

Equations (4.17) to (4.19) represent the first-order to third-order spatial partitioning of input-output linkages. So equation (4.16) can be considered as the zero-order spatial partition. The direct, indirect, and total effects for each of these equations are defined in the same way as in the preview paragraph. In equation (4.17), for example, the direct effects are generated from rounds of input-output processes within the own industry from the first-order neighboring regions. The indirect effects are generated from other industries through input-output linkages from neighboring regions. The total effects include all input-output effects from those regions. As the order of spatial neighbors increases, the magnitude of these effects diminishes given that the absolute value of ρ is less than one and $\boldsymbol{W}$ satisfies some regularity conditions (Elhorst, 2010a).

The spatially-oriented partitioned effects are related with the summary measures defined in Table 4.1 in the following ways. The sum of all partitioned total effects in the infinite series expansion should equal to the total space-industry effect. The sum of all partitioned direct effects should equal to the direct space-industry effect. The sum of all partitioned indirect effects should equal to the sum of the direct within-industry effect and the indirect space-industry effect. Since it is complicated to attribute the diagonal block matrices in equations (4.15) to (4.19) to some within-industry effects, and also to make the definition of the partitioned effects in both spatial partition and IO partition consistent, I simply define only one type of the partitioned indirect effect.

IO-Oriented Partitioned Effects

The IO-oriented partition is derived analogously to the spatially-oriented one only with the expansion of $\boldsymbol{Q}$ re-arranged in a different way as follows,

$$
\begin{aligned}
\boldsymbol{Q} &= \left[(\boldsymbol{I}_S - \phi\boldsymbol{A}) \otimes (\boldsymbol{I}_N - \rho\boldsymbol{W})\right]^{-1} \\
&= \boldsymbol{I}_{NS} + (\phi\boldsymbol{A} + \phi^2\boldsymbol{A}^2 + \phi^3\boldsymbol{A}^3 + \cdots) \otimes \boldsymbol{I}_N && (4.20) \\
&+ \boldsymbol{I}_S \otimes (\rho\boldsymbol{W} + \rho^2\boldsymbol{W}^2 + \rho^3\boldsymbol{W}^3 + \cdots) && (4.21) \\
&+ \phi\boldsymbol{A} \otimes (\rho\boldsymbol{W} + \rho^2\boldsymbol{W}^2 + \rho^3\boldsymbol{W}^3 + \cdots) && (4.22) \\
&+ \phi^2\boldsymbol{A}^2 \otimes (\rho\boldsymbol{W} + \rho^2\boldsymbol{W}^2 + \rho^3\boldsymbol{W}^3 + \cdots) && (4.23) \\
&+ \phi^3\boldsymbol{A}^3 \otimes (\rho\boldsymbol{W} + \rho^2\boldsymbol{W}^2 + \rho^3\boldsymbol{W}^3 + \cdots) && (4.24) \\
&+ \cdots\cdots
\end{aligned}
$$

Equation (4.20) encompasses all input-output multiplier effects, including the direct inputs before the first-round input-output process. Equation (4.21) represents the pure spatial spillover effects from other regions without a round of input-output processes. Equations (4.22) to (4.24) capture the spatial spillover effects from other regions with the first to third rounds of input-output processes. The direct, indirect, and total effects for each equation are defined analogously to those in the spatially-oriented partition. As the production moves on with many rounds of input-output processes, the partitioned effects estimates are expected to diminish given that the absolute value of ϕ is less than one and $\boldsymbol{A}$ satisfies some regularity conditions (Miller and Blair, 2009). Also, the relationship between the IO-oriented partitioned effects and the summary measures is maintained as in the spatially-oriented partition.

4.4 Monte Carlo Results

I evaluate the performance of the Bayesian MCMC estimation using Monte Carlo experiments. Each simulation randomly generates a row-standardized contiguity-based spatial weight matrix ($\boldsymbol{W}$), an input-output matrix ($\boldsymbol{A}$), two explanatory variables ($\boldsymbol{X}$), a set of industry specific effects ($\boldsymbol{\eta}$), and disturbances ($\boldsymbol{\varepsilon}$). To make the simulated sample mimic the actual data, one explanatory variable varies in both regions and industries, and the other explanatory variable only varies in industries so that it can be considered as some regional-specific factor. The dependent variable is then computed according to equation (4.2). For each simulation case, I use 100,000 iterations after a burn-in period of 50,000.

Tables 4.2-4.3 present the results of Monte Carlo experiments. Each table contains the true value of parameters in the second column, and the posterior mean, standard deviation, the 5% quantile, and the 95% quantile of parameter estimates. I draw ρ and ϕ both jointly and conditionally according to equations (4.33) and (4.34) in the appendix. I use two sample sizes, one with 50 regions and 20 industries, and another with 100 regions and 50 industries. As the sample size increases, I expect to see more accurate parameter estimates. As shown in these two tables, the sample means of all parameters are close to their true values. Drawing ρ and ϕ either jointly or conditionally does not make noticeable difference in their estimates. When the sample size increases, the estimates get closer to their true values, and the standard deviations get smaller.

Table 4.4 reports the true effects, the estimation mean, standard deviation, the 5% quantile, and the 95% quantile of all effect estimates defined in Table 4.1 with the sample of 100 regions and 50 industries and ρ and ϕ drawn conditionally. As shown, the total

Table 4.2: Monte Carlo Results for $N = 50$ and $S = 20$

parameter	true value	mean	s.d.	5%	95%	mean	s.d.	5%	95%
		draw ρ and ϕ jointly				draw ρ and ϕ conditionally			
β_1	0.6368	0.5917	0.0681	0.4797	0.7037	0.5917	0.0681	0.4805	0.7035
β_2	0.7930	0.7749	0.0960	0.6183	0.9345	0.7697	0.0966	0.6117	0.9301
σ_ε	0.2156	0.2200	0.0103	0.2033	0.2373	0.2203	0.0103	0.2036	0.2376
σ_η	0.4796	0.4978	0.1925	0.2433	0.8546	0.4976	0.1911	0.2444	0.8525
ρ	0.5383	0.5078	0.0414	0.4397	0.5764	0.5081	0.0405	0.4399	0.5739
ϕ	0.1969	0.2063	0.0798	0.0740	0.3366	0.2114	0.0805	0.0771	0.3421

Table 4.3: Monte Carlo Results for $N = 100$ and $S = 50$

parameter	true value	mean	s.d.	5%	95%	mean	s.d.	5%	95%
		draw ρ and ϕ jointly				draw ρ and ϕ conditionally			
β_1	0.3923	0.3942	0.0087	0.3799	0.4085	0.3945	0.0088	0.3801	0.4090
β_2	0.0447	0.0461	0.0092	0.0309	0.0611	0.0475	0.0095	0.0320	0.0631
σ_ε	2.7557	2.7666	0.0577	2.6723	2.8622	2.7625	0.0577	2.6689	2.8582
σ_η	0.4481	0.3137	0.0800	0.2001	0.4587	0.3098	0.0790	0.1969	0.4545
ρ	0.9182	0.9198	0.0059	0.9102	0.9299	0.9192	0.0061	0.9093	0.9297
ϕ	0.4271	0.4089	0.0283	0.3701	0.4619	0.3912	0.0396	0.3262	0.4561

space-industry effects of two parameters are both about 13 times greater than the direct space-industry effects, displaying strong spatial and inter-industry linkages in the simulated sample. The magnitude of effects estimates depends on the spatial weight matrix and the input-output matrix as well as ρ and ϕ.

Table 4.5 reports the spatially-oriented and IO-oriented partitioned effects for β_1. The first row in the left panel corresponds to equation (4.15) and the rest rows correspond to equations (4.16) to (4.19) and higher orders of neighbors. The first row in the right panel corresponds to equation (4.20), and the rest rows correspond to equations (4.21) to (4.24) and more rounds of input-output processes. With the simulated sample and the estimated parameters, the total pure spatial spillover effects are greater than the total pure input-output effects. In the spatially-oriented partition case, the total and indirect effects do not diminish to zero with the ninth-order neighbors, implying strong input-output linkages between industries over a wide spatial area, which is reasonable given that $\rho = 0.92$. The

Table 4.4: Effect Estimates for a Sample with $N = 100$ and $S = 50$

effects	parameters	true effects	mean	s.d	5% quantile	95% quantile
total space-industry	β_1	8.2619	8.0104	0.8082	6.7982	9.4362
effect	β_2	0.9404	0.9607	0.1962	0.6479	1.2906
direct space-industry	β_1	0.5851	0.5894	0.0151	0.5648	0.6148
effect	β_2	0.0666	0.0710	0.0141	0.0478	0.0944
direct within-industry	β_1	4.2763	4.3816	0.3729	3.8183	5.0721
effect	β_2	0.4868	0.5277	0.1141	0.3472	0.7198
indirect	β_1	3.4005	3.0395	0.5574	2.2049	4.0254
space-industry effect	β_2	0.3871	0.3621	0.0832	0.2366	0.5083

direct effect with the first-order neighbors is zero due to the diagonal elements of zeros in $\boldsymbol{W}$. The direct effects are close to zero after the second-order neighbors, implying that the own-industry input-output linkages are strong only in the own region. In the IO-oriented partition case, the total and indirect effects diminish faster than in the spatial partition case because the estimated ϕ is only 0.41. The direct effects also vanish quickly, implying that the feedback effects in the own region are small after a few rounds of input-output processes. Then, compare the sum of each column in the table and the last line containing the summary measures from Table 4.4. Since the partitioned direct effects diminish fast, the cumulative sum of them is very close to the summary direct effect, which is the direct space-industry effect. The cumulative sum of IO-oriented partitioned total and indirect effects are closer to their summary counterparts than of the spatially-oriented partitioned total and indirect effects because the speed of decay to zero in the IO-oriented partition is faster.

4.5 Application

I apply the space-industry filter to a model of regional industry employment. New economic geography (NEG) emphasizes that the allocation of industry employment is determined by centripetal forces and centrifugal forces (Krugman, 1998). The centripetal forces mainly stem from the strength of final demand and inter-industry linkages in the local market. The

Table 4.5: The Partitioned Effect for β_1 in a Sample with $N = 100$ and $S = 50$

	Spatial Partition				Industrial Partition		
	Total	Direct	Indirect		Total	Direct	Indirect
Pure spatial	4.8168	0.5811	4.2358	Pure I-O	0.6409	0.3993	0.2416
Order of neighbors	Total	Direct	Indirect	Round of I-O	Total	Direct	Indirect
W^0	0.2464	0.0048	0.2416	A^0	4.4223	0.1865	4.2358
W^1	0.2265	0.0000	0.2265	A^1	1.7002	0.0014	1.6989
W^2	0.2082	0.0007	0.2075	A^2	0.6537	0.0006	0.6532
W^3	0.1914	0.0002	0.1912	A^3	0.2514	0.0002	0.2511
W^4	0.1759	0.0003	0.1757	A^4	0.0966	0.0001	0.0966
W^5	0.1617	0.0002	0.1616	A^5	0.0372	0.0000	0.0371
W^6	0.1487	0.0001	0.1485	A^6	0.0143	0.0000	0.0143
W^7	0.1366	0.0001	0.1365	A^7	0.0055	0.0000	0.0055
W^8	0.1256	0.0001	0.1255	A^8	0.0021	0.0000	0.0021
W^9	0.1155	0.0001	0.1154	A^9	0.0008	0.0000	0.0008
Sum	6.5534	0.5876	5.9658	Sum	7.8251	0.5881	7.2370

Summary Measures	Total Effect	Direct Effect	Indirect Effect
	8.0104	0.5894	7.4211

Notes: (1) The left panel contains the results of spatially-oriented partitioning, each row in which represents equations (4.15) to (4.19) and partitioning with the higher-order neighbors.

(2) The right panel contains the results of IO-oriented partitioning, each row in which represents equations (4.20) to (4.24) and partitioning with the more IO rounds.

(3) The row of sum contains the cumulative sum of all values above this row in each column.

(4) Summary measures are from Table 4.4. Total effect corresponds to total space-industry effect, direct effect corresponds to direct space-industry effect, and indirect effect corresponds to the sum of direct within-industry effect and indirect space-industry effect.

centrifugal forces come from intense competition and the congestion costs, such as high crime rate, and pollution. I set up a simple model to account for these centripetal and centrifugal forces in determination of industry employment in a region.

Since the space-industry filter incorporates the input-output linkages among industries, this model is mostly driven by final demand. It is analogous to the basic form of an input-output model, i.e., $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{f} = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{f}$, where $\boldsymbol{x}$ is an industry output vector and $\boldsymbol{f}$ is a final demand vector. I use the market potential variable to describe final demand, which is the sum of final demand for the products of an industry in a region and its surrounding regions. Final demand in the surrounding regions is calculated using the spatial weight matrix times the local final demand for each industry in those regions. This definition of the market potential variable conforms with the concept of nominal market potential used in Harris (1954).

I use three other variables as control variables along with market potential to regress on regional industry employment. The growth rate of population from 1980 to 2000 accounts for the demographic condition in a region. The share of population with a Bachelor's or high school degree in 2000 represents skilled labor in a region. And the number of violent crimes in a region divided by the national number in 2000 serves as a proxy for negative living conditions as a centrifugal force.

To save computation time, I limit the sample to a moderate size by choosing the spatial units and industry aggregation in the estimation. Spatial dependence may vary with different spatial units. Thus, I estimate the model at both the state and county levels, with 49 continental states including Washington D.C., and 65 counties in the New England Region. The reason for choosing the New England region is that geographic features in that region are not very different across counties. The number of industries is 56, including industries with two or three digit NAICS codes. (See Table 4.13 in the appendix for a list of industries.)

Also, I use two types of spatial weight matrices to test the robustness of results, with one being a five-nearest neighbor weight matrix and another being an inverse distance weight matrix. The inverse-distance weight matrix is constructed using the cut-off distance that is 1.5 times as long as the distance ensuring a county to have at least one neighbor. Both types of spatial weight matrices are row-standardized. Finally, the input-output matrix is computed and aggregated using BEA's 2002 Use and Make tables at the summary level, without row standardization. Table 4.6 shows the descriptive statistics of variables used in estimation at the state and county levels. The variables of regional industry employment and market potential are logarithmically transformed.

Table 4.6: Descriptive Statistics of Variables in a Regional Industry Employment Model

label	variables	unit	min	mean	median	max	s.d.
		State level data					
y	Employment	persons	0	49210	11734	2681506	127081
x_1	MarketPotential	thousand \$	0.0000	5154540	1682960	167987114	10365743
x_2	PopGrowth	percentage	-0.4248	0.9976	0.5836	6.1954	1.0851
x_3	PercentBachelor	percentage	0.1483	0.2402	0.2324	0.3907	0.0479
x_4	Crime	percentage	0.0000	2.0241	1.0011	15.5972	2.8231
		County level data: New England Region					
y	Employment	persons	0	2006	226	99142	6546
x_1	MarketPotential	thousand \$	0.0000	239	40	12909	680
x_2	PopGrowth	percentage	-0.7787	0.6980	0.5305	2.3299	0.5868
x_3	PercentBachelor	percentage	0.1079	0.2633	0.2661	0.4359	0.0767
x_4	Crime	percentage	0.0000	0.0564	0.0073	0.6193	0.1120

Notes: (1) At the state level, the total number of observations is 2744, with 49 states (including DC) and 56 industries.

(2) At the county level, the total number of observations is 3640, with 65 counties and 56 industries.

(3) Data sources: County Business Patterns, USA Counties^{TM} and the cartographic boundary files from the Census Bureau.

(4) In estimation, y and x_1 is logarithmically transformed.

Tables 4.7 and 4.8 present the estimation results at the state level and the county level, respectively. The results using the five-nearest neighbor weight matrix are in the left panel and those using the inverse distance weight matrix are in the right panel. In these two tables, the parameters of interest are ρ and ϕ. As shown in the tables, ρ and ϕ are significantly positive at both the state level and the county level. The estimated ρ is significantly higher at the county level than at the state level. It is reasonable because at the county level, as the area of a region gets smaller, the spatial connection between regions becomes more substantial. At both the state and county levels, ϕ is greater than ρ, which implies that the inter-industry dependence plays a more important role than inter-regional dependence in the spatial distribution of industry employment.

Table 4.7: Estimation Results at the State Level of the Regional Industry Employment Model

parameter	mean	s.d.	5%	95%	mean	s.d.	5%	95%
	5-nearest neighbor weight matrix				inverse-distance weight matrix			
β_1	0.1098	0.0193	0.0800	0.1431	0.1112	0.0193	0.0807	0.1440
β_2	0.0141	0.0224	-0.0228	0.0507	0.0160	0.0224	-0.0211	0.0528
β_3	0.1404	0.5291	-0.7350	1.0046	0.2148	0.5221	-0.6393	1.0731
β_4	0.0634	0.0097	0.0475	0.0794	0.0646	0.0097	0.0487	0.0806
σ_ε	0.6465	0.0187	0.6161	0.6776	0.6434	0.0188	0.6129	0.6746
σ_η	0.3943	0.1011	0.2488	0.5769	0.3907	0.1000	0.2481	0.5721
ρ	0.1804	0.0301	0.1292	0.2284	0.1790	0.0313	0.1252	0.2299
ϕ	0.7290	0.0171	0.6998	0.7565	0.7254	0.0173	0.6972	0.7535

Table 4.8: Estimation Results for Counties in New England Region of the Regional Industry Employment Model

parameter	mean	s.d.	5%	95%	mean	s.d.	5%	95%
	5-nearest neighbor weight matrix				inverse-distance weight matrix			
β_1	0.1324	0.0300	0.0844	0.1827	0.1082	0.0263	0.0659	0.1522
β_2	-0.3211	0.1532	-0.5730	-0.0694	-0.2779	0.1541	-0.5299	-0.0242
β_3	4.2347	1.0418	2.5294	5.9687	3.7811	1.0026	2.1466	5.4160
β_4	2.8848	0.8369	1.5060	4.2579	3.0277	0.8206	1.6878	4.3859
σ_ε	0.0405	0.0011	0.0387	0.0423	0.0405	0.0011	0.0387	0.0422
σ_η	0.1079	0.0231	0.0734	0.1491	0.1524	0.0363	0.1000	0.2179
ρ	0.2512	0.0251	0.2098	0.2930	0.3701	0.0345	0.3128	0.4266
ϕ	0.6717	0.0198	0.6388	0.7041	0.6739	0.0192	0.6425	0.7057

Tables 4.9 and 4.10 present the effects estimates of the model at the state and county levels, respectively. Let's focus on the estimated effects of market potential, which is β_1 in the tables. It has significantly positive effects on regional industry employment. At both the state and county levels, the estimated coefficients on market potential are around 0.10-0.13 in Tables 4.7 and 4.8. According to the definition, the total space-industry effect of market

potential on regional industry employment is 0.49 at the state level and 0.54 at the county level. That means a one percent increase in final demand for an industry in a region will increase employment of all industries in all regions by about 0.5%. At the state level, a one percent increase in final demand for an industry is approximately equivalent to $51 million at the average in the sample, and a 0.5% increase in industry employment is approximately equivalent to 246 employees. The direct space-industry effects are 0.14 and 0.16 at the state and county levels, respectively, implying that a one percent increase in market potential for an industry in a region will increase employment within the same industry and the same region by 0.15%. The direct within-industry effects are 0.03 and 0.05 at the state and county levels, respectively, which are smaller than the direct space-industry effects. This shows that the market size effects emphasized in New Economic Geography are at work, reflecting the impacts of local demand on industrial location. Finally, the indirect space-industry effects are around 0.33 at both the state and county levels. The inter-dependence between different industries and neighboring regions is so high that it can account for more than half of the total effects. It implies the existence of agglomeration economies of a variety of industries in a broad geographic space.

Tables 4.11 and 4.12 show spatially- and IO-oriented partitioned effects for market potential, β_1, at the state and county levels. At both the state and levels, in contrast with the simulated sample, the spatially partitioned total and indirect effects decay at a higher speed than the IO partitioned effects. It implies that a shock in market potential for an industry may trigger many rounds of input-output processes in other related industries, while its impacts on spatial neighbors may not be too wide. Both spatially- and IO-oriented partitioned direct effects vanish after at most the second round. Except for the pure spatial and pure IO effects in the first line in the tables, the magnitude of direct effects after the second row are dominated by the indirect effects, implying that market potential for an industry in a region may have stronger impacts on related industries in neighboring regions than impacts on its own industry in the own region.

Table 4.9: Effect Estimates at the State Level with a Five-Nearest Neighbor Weight Matrix

effect	parameter	mean	s.d	5%	95%
total space-industry effect	β_1	0.4936	0.0726	0.3770	0.6155
	β_2	0.0638	0.1020	-0.1030	0.2303
	β_3	0.6872	2.4109	-3.2142	4.7145
	β_4	0.2861	0.0425	0.2167	0.3568
direct space-industry effect	β_1	0.1376	0.0235	0.1010	0.1780
	β_2	0.0177	0.0281	-0.0286	0.0636
	β_3	0.1777	0.6636	-0.9180	1.2657
	β_4	0.0795	0.0119	0.0599	0.0990
direct within-industry effect	β_1	0.0292	0.0064	0.0194	0.0403
	β_2	0.0038	0.0062	-0.0061	0.0143
	β_3	0.0429	0.1464	-0.1879	0.2913
	β_4	0.0171	0.0044	0.0104	0.0248
indirect space-industry effect	β_1	0.3268	0.0479	0.2501	0.4074
	β_2	0.0424	0.0679	-0.0684	0.1534
	β_3	0.4666	1.6054	-2.1036	3.1667
	β_4	0.1895	0.0288	0.1432	0.2381

Table 4.10: Effect Estimates with New England Region with a Five-Nearest Neighbor Weight Matrix

effects	parameter	mean	s.d	5%	95%
total space-industry effect	β_1	0.5394	0.1201	0.3471	0.7405
	β_2	-1.3100	0.6282	-2.3513	-0.2865
	β_3	17.2398	4.1426	10.4163	24.0879
	β_4	11.7336	3.3175	6.2437	17.1443
direct space-industry effect	β_1	0.1622	0.0365	0.1036	0.2234
	β_2	-0.3936	0.1876	-0.7015	-0.0853
	β_3	5.1887	1.2667	3.1093	7.2917
	β_4	3.5342	1.0180	1.8511	5.2022
direct within-industry effect	β_1	0.0519	0.0125	0.0323	0.0732
	β_2	-0.1268	0.0634	-0.2353	-0.0268
	β_3	1.6654	0.4479	0.9618	2.4297
	β_4	1.1358	0.3568	0.5755	1.7435
indirect space-industry effect	β_1	0.3252	0.0748	0.2071	0.4520
	β_2	-0.7896	0.3821	-1.4271	-0.1722
	β_3	10.3857	2.5361	6.2716	14.6115
	β_4	7.0637	2.0063	3.7667	10.3453

Table 4.11: The Partitioned Effects for β_1 at the State Level

	Spatial Partition				Industrial Partition		
	Total	Direct	Indirect		Total	Direct	Indirect
Pure spatial	0.1340	0.1105	0.0236	Pure I-O	0.4053	0.1369	0.2684
Order of neighbors	Total	Direct	Indirect	Round of I-O	Total	Direct	Indirect
W^0	0.2955	0.0271	0.2684	A^0	0.0242	0.0006	0.0236
W^1	0.0533	0.0000	0.0533	A^1	0.0176	0.0001	0.0175
W^2	0.0096	0.0001	0.0095	A^2	0.0128	0.0000	0.0128
W^3	0.0017	0.0000	0.0017	A^3	0.0094	0.0000	0.0094
W^4	0.0003	0.0000	0.0003	A^4	0.0068	0.0000	0.0068
W^5	0.0001	0.0000	0.0001	A^5	0.0050	0.0000	0.0050
W^6	0.0000	0.0000	0.0000	A^6	0.0036	0.0000	0.0036
W^7	0.0000	0.0000	0.0000	A^7	0.0026	0.0000	0.0026
W^8	0.0000	0.0000	0.0000	A^8	0.0019	0.0000	0.0019
W^9	0.0000	0.0000	0.0000	A^9	0.0014	0.0000	0.0014
Sum	0.4945	0.1377	0.3568	Sum	0.4907	0.1377	0.3531
Summary Measures	Total Effect		Direct Effect		Indirect Effect		
	0.4936		0.1376		0.3560		

Notes: (1) The left panel contains the results of spatially-oriented partitioning, each row in which represents equations (4.15) to (4.19) and partitioning with the higher-order neighbors.

(2) The right panel contains the results of IO-oriented partitioning, each row in which represents equations (4.20) to (4.24) and partitioning with the more IO rounds.

(3) The row of sum contains the cumulative sum of all values above this row in each column.

(4) Summary measures are from Table 4.9. Total effect corresponds to total space-industry effect, direct effect corresponds to direct space-industry effect, and indirect effect corresponds to the sum of direct within-industry effect and indirect space-industry effect.

Table 4.12: The Partitioned Effects for β_1 in New England Counties

	Spatial Partition				Industrial Partition		
	Total	Direct	Indirect		Total	Direct	Indirect
Pure spatial	0.1768	0.1339	0.0429	Pure I-O	0.4033	0.1604	0.2429
Order of neighbors	Total	Direct	Indirect	Round of I-O	Total	Direct	Indirect
W^0	0.2709	0.0280	0.2429	A^0	0.0444	0.0015	0.0429
W^1	0.0681	0.0000	0.0681	A^1	0.0298	0.0002	0.0296
W^2	0.0171	0.0003	0.0168	A^2	0.0200	0.0001	0.0200
W^3	0.0043	0.0000	0.0043	A^3	0.0135	0.0000	0.0134
W^4	0.0011	0.0000	0.0011	A^4	0.0090	0.0000	0.0090
W^5	0.0003	0.0000	0.0003	A^5	0.0061	0.0000	0.0061
W^6	0.0001	0.0000	0.0001	A^6	0.0041	0.0000	0.0041
W^7	0.0000	0.0000	0.0000	A^7	0.0027	0.0000	0.0027
W^8	0.0000	0.0000	0.0000	A^8	0.0018	0.0000	0.0018
W^9	0.0000	0.0000	0.0000	A^9	0.0012	0.0000	0.0012
Sum	0.5386	0.1622	0.3764	Sum	0.5361	0.1622	0.3738

Summary Measures	Total Effect	Direct Effect	Indirect Effect
	0.5394	0.1622	0.3771

Notes: (1) The left panel contains the results of spatially-oriented partitioning, each row in which represents equations (4.15) to (4.19) and partitioning with the higher-order neighbors.

(2) The right panel contains the results of IO-oriented partitioning, each row in which represents equations (4.20) to (4.24) and partitioning with the more IO rounds.

(3) The row of sum contains the cumulative sum of all values above this row in each column.

(4) Summary measures are from Table 4.10. Total effect corresponds to total space-industry effect, direct effect corresponds to direct space-industry effect, and indirect effect corresponds to the sum of direct within-industry effect and indirect space-industry effect.

4.6 Conclusion

In this paper, I explore the space-industry filter to model regional industry activity. The space-industry filter incorporates both inter-regional and inter-industry relationships in one setting, extending the literature of integrating spatial econometrics and the input-output model. The approaches to interpreting the effects estimates and partitioning by order of spatial neighbors and round of input-output processes provide us useful and detailed information about the strength of spatial spillovers and input-output linkages. The results from Monte Carlo experiments and an application of a regional industry employment model confirm the feasibility of this modeling technique. In the regional industry employment model, market potential is shown to be an important centripetal factor in industry location decisions.

The space-industry filter has wide applicability. One potential direction is to examine the impacts of regional development policies. To attract firms to invest in a region, local governments often offer firms tax benefits or relax regulations. Using the space-industry filter, we can examine the multiplier effects of these policies, evaluating how much employment and income these policies can create from not only the intended industries but also from all related industries. Moreover, we can investigate how the beneficial polices in neighboring regions can affect industrial structure in the local economy. Future research will also explore new ways to build a model with the space-industry filter and dig deeper into interpretation of effects estimates.

Appendix

The Derivation of the Conditional Posterior Distributions

From the joint posterior distribution (4.6)-(4.10), we get the kernel of the conditional posterior distribution of h_ε is

$$f(h_\varepsilon|\boldsymbol{\beta}, \boldsymbol{\eta}, h_\eta, \rho, \phi) \propto h_\varepsilon^{\frac{v_0+NS}{2}-1} \exp\left(-\frac{\delta_0 + \boldsymbol{e}'\boldsymbol{e}}{2} h_\varepsilon\right) \tag{4.25}$$

which has the form of a Gamma distribution, i.e.,

$$h_\varepsilon|\boldsymbol{\beta}, \boldsymbol{\eta}, \rho, \phi, \boldsymbol{y} \sim G(\bar{v}_0/2, \bar{\delta}_0/2) \tag{4.26}$$

where $\bar{v}_0 = v_0 + NS$ and $\bar{\delta}_0 = \delta_0 + \boldsymbol{e}'\boldsymbol{e}$.

Similarly, the conditional posterior distribution of h_η is also a Gamma distribution as,

$$h_\eta|\boldsymbol{\eta}, \boldsymbol{y} \sim G(\bar{v}_1/2, \bar{\delta}_1/2) \tag{4.27}$$

where $\bar{v}_1 = v_1 + S$ and $\bar{\delta}_1 = \delta_1 + \boldsymbol{\eta}'\boldsymbol{\eta}$.

Due to the concern of possible correlation, $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are sampled in one block as follows

$$f(\boldsymbol{\beta}, \boldsymbol{\eta}|h_\varepsilon, h_\eta, \rho, \phi, \boldsymbol{y}) = f(\boldsymbol{\beta}|h_\varepsilon, h_\eta, \rho, \phi, \boldsymbol{y}) f(\boldsymbol{\eta}|\boldsymbol{\beta}, h_\varepsilon, h_\eta, \rho, \phi, \boldsymbol{y}) \tag{4.28}$$

The first term on the right-hand side can be found by integrating out $\boldsymbol{\eta}$ from the term on the left-hand side. For the second term, from the posterior distribution (4.28), the conditional posterior distribution of $\boldsymbol{\eta}$ is given by

$$\boldsymbol{\eta}|\boldsymbol{\beta}, h_\varepsilon, h_\eta, \rho, \phi, \boldsymbol{y} \sim N(\boldsymbol{\mu}_1, \boldsymbol{D}_1) \tag{4.29}$$

where $\boldsymbol{D}_1 = h_\varepsilon^{-1}/(N + \frac{h_\eta}{h_\varepsilon})\boldsymbol{I}_S$, $\boldsymbol{\mu_1} = 1/(N + \frac{h_\eta}{h_\varepsilon})\boldsymbol{E}'\tilde{\boldsymbol{y}}$, and $\tilde{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$.

Integrating out $\boldsymbol{\eta}$, the conditional posterior distribution of $\boldsymbol{\beta}$ is given by

$$\begin{aligned} f(\boldsymbol{\beta}|h_\varepsilon, h_\eta, \rho, \phi, \boldsymbol{y}) &\propto \exp\left(-\frac{1}{2}\tilde{\boldsymbol{y}}'\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{y}}\right) \\ &\quad \times \exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{M}_\beta(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right) \\ &\propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_1)'(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}+\boldsymbol{M}_\beta)(\boldsymbol{\beta}-\boldsymbol{\beta}_1)\right) \end{aligned} \tag{4.30}$$

Combining terms in equation (4.30), $\boldsymbol{\beta}$ can be seen as normally distributed, i.e.,

$$\boldsymbol{\beta}|h_\varepsilon, h_\eta, \rho, \phi, \boldsymbol{y} \sim N(\boldsymbol{\beta}_1, \boldsymbol{D}_2) \tag{4.31}$$

where

$$\begin{aligned} \boldsymbol{D}_2 &= \left(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}+\boldsymbol{M}_\beta\right)^{-1} \\ \boldsymbol{\beta}_1 &= \boldsymbol{D}_2\left(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}\boldsymbol{y}+\boldsymbol{M}_\beta\boldsymbol{\beta}_0\right) \\ \boldsymbol{\Omega} &= h_\eta^{-1}(\boldsymbol{I}_S\otimes\boldsymbol{\iota}_N\boldsymbol{\iota}_N') + h_\varepsilon^{-1}\boldsymbol{I}_{NS} \end{aligned}$$

Finally, we use the Metropolis-Hasting (MH) algorithm to draw samples of ρ and ϕ because the conditional posterior distributions of these two parameters are not in the form of a familiar distribution. The joint conditional posterior distribution for them is given by

$$f(\rho, \phi|\boldsymbol{\beta}, h_\varepsilon, h_\eta, \phi, \boldsymbol{y}) \propto |\boldsymbol{P}(\rho,\phi)| \exp\left(-\frac{h_\varepsilon}{2}\boldsymbol{e}'(\rho,\phi)\boldsymbol{e}(\rho,\phi)\right) \tag{4.32}$$

Following Parent and LeSage (2012), the transitional kernel for ρ and ϕ is a random walk process with the normally distributed disturbance, i.e., $\rho^{(j+1)} = \rho^{(j)} + c_1N(0,1)$ and $\phi^{(j+1)} = \phi^{(j)} + c_2N(0,1)$. c_1 and c_2 are the tuning parameters, as suggested in LeSage and Pace (2010), for ensuring that the MH sampling procedure moves over the entire conditional distribution. The proposed draws of ρ and ϕ is confined in the interval of $(-1,1)$. All draws with ρ and ϕ being outside this interval are rejected. The proposed draws of ρ and ϕ within the interval are then used to compute the acceptance probability. When we draw ρ and ϕ jointly from their joint posterior distribution (equation 4.32), the proposed samples are rejected based on the acceptance probability, given by

$$\alpha\left((\rho^{(j)},\phi^{(j)}),(\rho^{(j+1)},\phi^{(j+1)}\right) = \min\left\{1, \frac{f(\rho^{(j+1)},\phi^{(j+1)}|\boldsymbol{\beta},h_\varepsilon,h_\eta,\phi,\boldsymbol{y})}{f(\rho^{(j)},\phi^{(j)}|\boldsymbol{\beta},h_\varepsilon,h_\eta,\phi,\boldsymbol{y})}\right\} \tag{4.33}$$

We can also draw ρ and ϕ separately using the idea of $f(\rho,\phi) = f(\rho)f(\phi|\rho)$. A new sample of ϕ is drawn first conditional on the current value of ρ, and then ρ is drawn using the new ϕ. The acceptance probability for each parameter is

$$\begin{aligned}\alpha(\phi^{(j)},\phi^{(j+1)}) &= \min\left\{1, \frac{f(\phi^{(j+1)}|\rho^{(j)},\boldsymbol{\beta},h_\varepsilon,h_\eta,\boldsymbol{y})}{f(\phi^{(j)}|\rho^{(j)},\boldsymbol{\beta},h_\varepsilon,h_\eta,\boldsymbol{y})}\right\}\\ \alpha(\rho^{(j)},\rho^{(j+1)}) &= \min\left\{1, \frac{f(\rho^{(j+1)}|\phi^{(j+1)},\boldsymbol{\beta},h_\varepsilon,h_\eta,\boldsymbol{y})}{f(\rho^{(j)}|\phi^{(j+1)},\boldsymbol{\beta},h_\varepsilon,h_\eta,\boldsymbol{y})}\right\}\end{aligned} \tag{4.34}$$

The conditional posterior distributions for ρ and ϕ take the same form as in equation (4.33) because they can not be integrated out.

The Bayesian MCMC Algorithm

In summary, the algorithm of the Bayesian MCMC method is the following,

Step 1 Choose $\boldsymbol{\beta}^{(0)},\boldsymbol{\eta}^{(0)},\rho^{(0)},\phi^{(0)}$

Step 2 At the jth iteration, use the Gibbs algorithm to sample

$$\begin{aligned} h_\varepsilon^{(j+1)} &\sim G(\bar{v}_0/2,\bar{\delta}_0^{(j)}/2)\\ h_\eta^{(j+1)} &\sim G(\bar{v}_1/2,\bar{\delta}_1^{(j)}/2)\\ \boldsymbol{\eta}^{(j+1)} &\sim N(\boldsymbol{\mu}_1^{(j)},\boldsymbol{D}_1^{(j)})\\ \boldsymbol{\beta}^{(j+1)} &\sim N(\boldsymbol{\beta}_1^{(j)},\boldsymbol{D}_2^{(j)})\end{aligned}$$

where

$$
\begin{aligned}
\bar{\delta}_0^{(j)} &= \delta_0 + (\boldsymbol{e}^{(j)})'\boldsymbol{e}^{(j)} \\
\boldsymbol{e}^{(j)} &= \boldsymbol{P}^{(j)}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(j)} - \boldsymbol{E}\boldsymbol{\eta}^{(j)} \\
\boldsymbol{P}^{(j)} &= (\boldsymbol{I}_S - \phi^{(j)}\boldsymbol{A}) \otimes (\boldsymbol{I}_N - \rho^{(j)}\boldsymbol{W}) \\
\bar{\delta}_1^{(j)} &= \delta_1 + (\boldsymbol{\eta}^{(j)})'\boldsymbol{\eta}^{(j)} \\
\boldsymbol{D}_1^{(j)} &= \frac{(h_\varepsilon^{(j)})^{-1}}{N + h_\eta^{(j)}/h_\varepsilon^{(j)}}\boldsymbol{I}_S, \\
\boldsymbol{\mu_1}^{(j)} &= \frac{1}{N + h_\eta^{(j)}/h_\varepsilon^{(j)}}\boldsymbol{E}'(\boldsymbol{P}^{(j)}\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(j)}) \\
\boldsymbol{D}_2^{(j)} &= \left(\boldsymbol{X}'(\boldsymbol{\Omega}^{(j)})^{-1}\boldsymbol{X} + \boldsymbol{M}_\beta\right)^{-1} \\
\boldsymbol{\beta}_1^{(j)} &= \boldsymbol{D}_2^{(j)}\left(\boldsymbol{X}'(\boldsymbol{\Omega}^{(j)})^{-1}\boldsymbol{P}^{(j)}\boldsymbol{y} + \boldsymbol{M}_\beta\boldsymbol{\beta}_0\right) \\
\boldsymbol{\Omega}^{(j)} &= (h_\eta^{(j)})^{-1}(\boldsymbol{I}_S \otimes \boldsymbol{\iota}_N\boldsymbol{\iota}_N') + (h_\varepsilon^{(j)})^{-1}\boldsymbol{I}_{NS}
\end{aligned}
$$

Step 3a (joint drawing) At the jth iteration, use the MH algorithm to draw $\rho^{(j+1)}$ and $\phi^{(j+1)}$ jointly from their joint posterior distribution.

1. Draw U from $U(0,1)$
2. Generate the proposed draw of $\rho^{(j+1)}$ and $\phi^{(j+1)}$ by $\rho^{(j+1)} = \rho^{(j)} + c_1 N(0,1)$ and $\phi^{(j+1)} = \phi^{(j)} + c_2 N(0,1)$. If $\rho^{(j+1)} \notin (-1,1)$ or $\phi^{(j+1)} \notin (-1,1)$, draw $\rho^{(j+1)}$ or $\phi^{(j+1)}$ again.
3. The acceptance probability is calculated as

$$
\alpha\left((\rho^{(j)}, \phi^{(j)}), (\rho^{(j+1)}, \phi^{(j+1)})\right) = \min\left\{1, \frac{|\boldsymbol{P}^{(j+1)}|\exp\left(-\frac{1}{2}h_\varepsilon^{(j+1)}(\boldsymbol{e}^{(j+1)})'\boldsymbol{e}^{(j+1)}\right)}{|\boldsymbol{P}^{(j)}|\exp\left(-\frac{1}{2}h_\varepsilon^{(j+1)}(\boldsymbol{e}^{(j)})'\boldsymbol{e}^{(j)}\right)}\right\}
$$

   where $\boldsymbol{e}^{(j)}$ is computed with $\boldsymbol{\beta}^{(j+1)}$ and $\boldsymbol{\eta}^{(j+1)}$.
4. if $U < \alpha\left((\rho^{(j)}, \phi^{(j)}), (\rho^{(j+1)}, \phi^{(j+1)})\right)$, accept $\rho^{(j+1)}$ and $\phi^{(j+1)}$. Otherwise, feed $\rho^{(j)}$ and $\phi^{(j)}$ in the next iteration.

Step 3b (conditional drawing) At the jth iteration, use the MH algorithm to sample

$\rho^{(j+1)}$ and $\phi^{(j+1)}$ separately.

1. Draw U_1 and U_2 from $U(0,1)$
2. Generate $\rho^{(j+1)}$ and $\phi^{(j+1)}$ by $\rho^{(j+1)} = \rho^{(j)} + c_1 N(0,1)$ and $\phi^{(j+1)} = \phi^{(j)} + c_2 N(0,1)$. If $\rho^{(j+1)} \notin (-1,1)$ or $\phi^{(j+1)} \notin (-1,1)$, draw $\rho^{(j+1)}$ or $\phi^{(j+1)}$ again.
3. if $U_1 \leq \alpha(\rho^{(j)}, \rho^{(j+1)})$, accept $\rho^{(j+1)}$. Otherwise, set $\rho^{(j+1)} = \rho^{(j)}$. Likewise, if $U_2 \leq \alpha(\phi^{(j)}, \phi^{(j+1)})$, accept $\phi^{(j+1)}$. Otherwise, set $\phi^{(j+1)} = \phi^{(j)}$
4. if $\alpha(\cdot) < 40\%$, set $c_{1,2} = c_{1,2}/1.1$. if $\alpha(\cdot) > 60\%$, set $c_{1,2} = 1.1c_{1,2}$.

Step 4 Repeat steps 1-3 until the desired number of samples is obtained. For an iteration of G times, the first B samples can be discarded as the burn-in samples. Compute the sample mean, standard deviation, the 5% quantile, and the 95% quantile, using the $G - B$ samples.

Table 4.13: Description of Industries Used in the Model Estimation

NAICS	Industry Description
11	Forestry, fishing, hunting, and agriculture support
211	Oil & gas extraction
212	Mining (except oil & gas)
213	Mining support activities
221	Utilities
23	Construction
311	Food mfg
312	Beverage & tobacco product mfg
313	Textile mills
314	Textile product mills
315	Apparel manufacturing
316	Leather & allied product mfg
321	Wood product mfg
322	Paper mfg
323	Printing & related support activities
324	Petroleum & coal products mfg
325	Chemical mfg
326	Plastics & rubber products mfg
327	Nonmetallic mineral product mfg
331	Primary metal mfg
332	Fabricated metal product mfg
333	Machinery mfg
334	Computer & electronic product mfg
335	Electrical equip, appliance & component mfg
336	Transportation equipment mfg
337	Furniture & related product mfg
339	Miscellaneous mfg
42	Wholesale trade
44	Retail trade
48	Transportation & warehousing
492	Couriers & messengers
493	Warehousing & storage
511	Publishing industries
512	Motion picture & sound recording industries
513	Broadcasting & telecommunications

514	Information & data processing services
52	Finance & insurance
531	Real estate
532	Rental & leasing services
533	Lessors of other nonfinancial intangible asset
541	Professional, scientific & technical services
551	Management of companies & enterprises
561	Administrative & support services
562	Waste management & remediation services
611	Educational services
621	Ambulatory health care services
622	Hospitals
623	Nursing & residential care facilities
624	Social assistance
71	Arts, entertainment & recreation
721	Accommodation
722	Food services & drinking places
811	Repair & maintenance
812	Personal & laundry services
813	Religious, grantmaking, civic, prof & like organizations
92	Government

Chapter 5

Conclusion and Future Research

5.1 Summary

This dissertation research explores new methods in studying industrial agglomeration, emphasizing the role of input-output linkages in the formation of agglomeration. First, I propose a new technique to measure agglomeration of an industry in a region by applying a bootstrap method to the standardized location quotient. The feature of this new method is that it is free of any statistical assumption imposed on the measurement, making it easy to use. The method is applied in detecting the existence of agglomeration of manufacturing industries in U.S counties in Chapter 2. Subsequently, Chapters 3 and 4 are organized around a central point of this dissertation: the two-dimensional characteristic of industrial agglomeration. The spatial and industry dimensions in regard to industrial agglomeration motivate this research to integrate spatial econometrics and the input-output model, addressing both spatial spillovers and inter-industry linkages simultaneously.

I use two methods to integrate spatial econometrics and the input-output model. The first method is to use the input-output matrix to generate the intermediate demand variable as an explanatory variable describing the strength of input-output linkages. This variable is then included in a spatial econometric model with the dependent variable being the location quotient. This method is a direct application of the embedding strategy of the integrated econometrics and input-output model in the literature. The second method is to devise a space-industry filter, which is essentially a weight matrix composed of both a spatial weight matrix and an input-output matrix. This method can be considered as the expansion of the embedding strategy mentioned above. Instead of introducing the input-output model through an explanatory variable, an input-output matrix is imposed directly on the dependent variable by means of the space-industry filter, resulting in a spatial econometric type model. Another appealing feature of this method lies in the interpretation of the coefficients on explanatory variables, which stresses both spatial spillover effects in the spatial econometric models and the multiplier effects in the input-output model.

5.2 Future Research

I believe that the space-industry filter method is very promising with a wide range of applications. I list here at least three potential directions that future research can pursue.

The impact analysis of regional development policies. For the purpose of attracting firms to invest in a region, local governments often offer firms tax benefits or relax regulations. Using the space-industry filter, we can examine the multiplier effects of these policies, evaluating how much more employment and income these policies can create from not only the intended industries but also from all related industries. Also, we can investigate how the beneficial polices in neighboring regions can affect industrial structure in the local economy.

The regional reaction to a macroeconomic shock. Regions may have disparate reaction to a macroeconomic shock. For example, a report of the Bureau of Labor Statistics shows that, during the economic recession from 2007 to 2010, some states experienced serious economic decline, such as Michigan and California, while some states were affected modestly, like North Dakota and Nebraska. Moreover, states with a high proportion of manufacturing and construction industries were hit worst, but states with strong education- and health-related industries mitigated the negative shock with some increasing employment in these industries. With the space-industry filter, we can investigate the expansion path of a macroeconomic shock across regions as well as industries.

Industry transfers with integration of global economies. Globalization gives rise to international transfer of industrial production. For example, since joining the WTO in 2001, China has become the largest manufacturer in the world. However, in recent years, some manufacturing industries have started moving out of China to neighboring countries, such as Vietnam and the Philippines, due to the increasing labor costs in China. Globalization has a profound influence on a country's industrial structure, which in turn determines the pattern of international trade. This topic of international movement of industries also fits in the idea of the space-industry filter, for which we can estimate the multiplier effects of an industry moving into and out of a country on economic growth of this country and its neighboring countries.

Bibliography

Anselin, L., Gallo, J. L., Jayet, H., 2008. Spatial Panel Econometrics. In: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data. Springer Berlin Heidelberg, Ch. 19, pp. 625–660.

Arbia, G., Espa, G., Giuliani, D., Mazzitelli, A., Sep. 2010. Detecting the existence of spacetime clustering of firms. Regional Science and Urban Economics 40 (5), 311–323.

Baldwin, R. E., Forslid, R., Aug. 2000. The Core-Periphery Model and Endogenous Growth: Stabilizing and Destabilizing Integration. Economica 67 (267), 307–324.

Behrens, K., Thisse, J.-F., Jul. 2007. Regional economics: A new economic geography perspective. Regional Science and Urban Economics 37 (4), 457–465.

Bertinelli, L., Decrop, J., 2005. Geographical agglomeration: Ellison and Glaeser's index applied to the case of Belgian manufacturing industry.

Christaller, W., 1933. Central Places in Southern Germany. Fischer, Jena, Germany.

Combes, P.-P., Mayer, T., Thisse, J.-F., 2008. Economic geography: the integration of regions and nations. Princeton University Press.

Davidson, R., MacKinnon, J. G., 2004. Econometric Theory and Methods. Oxford University Press, New York Oxford.

Debarsy, N., Ertur, C., LeSage, J. P., Jan. 2012. Interpreting dynamic spacetime panel data models. Statistical Methodology 9 (1-2), 158–171.

Dumais, G., Ellison, G., Glaeser, E. L., 1997. Geographic Concentration as a Dynamic Process. Working Paper Series No.6270, National Bureau of Economic Research.

Duranton, G., Overman, H. G., 2005. Testing for localization using micro-geographic data. The Review of Economic Studies 72 (4), 1077–1106.

Elhorst, J. P., Mar. 2010a. Applied Spatial Econometrics: Raising the Bar. Spatial Economic Analysis 5 (1), 9–28.

Elhorst, J. P., 2010b. Spatial Panel Data Models. In: Fisher, M. M., Getis, A. (Eds.), Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Springer Berlin Heidelberg and New York, Ch. C.2, pp. 377–405.

Ellison, G., Glaeser, E. L., Oct. 1997. Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. Journal of Political Economy 105 (5), 889.

Ellison, G., Glaeser, E. L., Kerr, W. R., 2010. What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. American Economic Review 100 (June), 1195–1213.

Forslid, R., Ottaviano, G. I. P., 2003. An analytically solvable core-periphery model. Journal of Economic Geography 3 (3), 229 – 240.

Fujita, M., Krugman, P. R., Venables, A. J., 1999. The Spatial Economy: Cities, Regions and International trade. MIT Press, Cambridge, MA.

Fujita, M., Thisse, J.-F., 2002. Economics of agglomeration: cities, industrial location, and regional growth. Cambridge University Press, Cambridge, UK.

Glaeser, E. L., Kerr, W. R., 2009. Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain? Journal of Economics Management Strategy 18 (3), 623–663.

Guimarães, P., Figueiredo, O., Woodward, D., May 2009. Dartboard tests for the location quotient. Regional Science and Urban Economics 39 (3), 360–364.

Hanson, N. M., 1967. Development pole theory in a regional context. Kyklos 20 (4), 709–727.

Harris, C. D., 1954. The Market as a Factor in the Localization of Industry in the United States. Annals of the Association of American Geographers 44 (4), 315–348.

Head, K., Mayer, T., 2004. The empirics of agglomeration and trade. Handbook of Regional and Urban Economics 4 (04), 2609–2669.

Holmes, T. J., Stevens, J. J., 2004. Spatial distribution of economic activities in North America. In: Henderson, J. V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4. Elsevier, Ch. 63, pp. 2797–2843.

Hotelling, H., 1929. Stability in Competition. The Economic Journal 39 (153), 41–57.

Isserman, A. M., Westervelt, J., Jul. 2006. 1.5 Million Missing Numbers: Overcoming Employment Suppression in County Business Patterns Data. International Regional Science Review 29 (3), 311–335.

Jofre-Monseny, J., Marín-López, R., Viladecans-Marsal, E., Sep. 2011. The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms. Journal of Urban Economics 70 (2-3), 61–74.

Kelejian, H. H., Prucha, I. R., 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. Journal of Real Estate Finance and Economics 17 (1), 99–121.

Kelejian, H. H., Prucha, I. R., 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. International Economic Review 40 (2), 509–533.

Koopmans, T. C., 1957. Three essays on the state of economic science. New York : McGraw-Hill.

Kranich, J., Mar. 2011. Agglomeration, vertical specialization, and the strength of industrial linkages. Papers in Regional Science 90 (1), 159–178.

Krugman, P. R., Jan. 1991. Increasing Returns and Economic Geography. Journal of Political Economy 99 (3), 483–499.

Krugman, P. R., 1998. What's new about the new economic geography? Oxford Review of Economic Policy 14 (2), 7–17.

Krugman, P. R., Venables, A. J., 1995. Globalization and the Inequality of Nations. Quarterly Journal of Economics 110 (4), 857–880.

LeSage, J. P., Pace, R. K., 2010. Introduction to Spatial Econometrics. Chapman and Hall/CRC.

Lösch, A., 1938. The Nature of Economic Regions. Sourther Economic Journal 5 (1), 71–78.

Marcon, E., Puech, F., 2003. Evaluating the geographic concentration of industries using distance-based methods. Journal of Economic Geography 3 (4), 409–428.

Marcon, E., Puech, F., Nov. 2010. Measures of the geographic concentration of industries: improving distance-based methods. Journal of Economic Geography 10 (5), 745–762.

Marshall, A., 1890. Principle of economics. Macmillan, London, UK.

Martin, P., Ottaviano, G. I. P., Feb. 1999. Growing locations: Industry location in a model of endogenous growth. European Economic Review 43 (2), 281–302.

Maurel, F., Sédillot, B., Sep. 1999. A measure of the geographic concentration in french manufacturing industries. Regional Science and Urban Economics 29 (5), 575–604.

Miller, R. E., Blair, P. D., 2009. Input-output analysis: foundations and extensions. Cambridge University Press.

Millo, G., Piras, G., 2012. splm : Spatial Panel Data Models in R. Journal Of Statistical Software 47 (1).

Mion, G., 2004. Input-Output Linkages, Proximity to Final Demand and the Location of Manufacturing Industries. CORE Discussion Paper No. 2004/53.

Moghadam, K., Ballard, K. P., 1988. Small area modeling of the industrial sector (SAMIS): an integrated econometric - interindustry approach. Environment and Planning A 20 (5), 655–668.

Moineddin, R., Beyene, J., Boyle, E., 2003. On the Location Quotient Confidence Interval. Geographical Analysis 35 (3), 249–256.

Mutl, J., Pfaffermayr, M., Feb. 2011. The Hausman test in a Cliff and Ord panel model. The Econometrics Journal 14 (1), 48–76.

Nakamura, R., Paul, C. J. M., 2009. Measuring agglomeration. In: Capello, R., Nijkamp, P. (Eds.), Handbook of regional growth and development theories. Edward Elgar Publishing, Northampton, MA, Ch. 16, pp. 305–328.

O'Donoghue, D., Gleave, B., 2004. A Note on Methods for Measuring Industrial Agglomeration. Regional Studies 38 (4), 419–427.

Openshaw, S., 1981. The modifiable areal unit problem. In: Wrigley, N., Bennett, R. J. (Eds.), Quantitative geography a British view. Vol. 38 of Concepts and techniques in modern geography,. Geo Books, Ch. 5, pp. 60–70.

O'Sullivan, A., 2011. Urban Economics, 8th Edition. Vol. 1. McGraw-Hill/Irwin.
URL `http://www.worldcat.org/title/urban-economics/oclc/49283880&referer=brief_results`

Ottaviano, G. I. P., Robert-Nicoud, F., 2006. The 'genome' of NEG models with vertical linkages: a positive and normative synthesis. Journal of Economic Geography 6 (2), 113–139.

Overman, H. G., Puga, D., 2010. Labor Pooling as a Source of Agglomeration: An Empirical Investigation. In: Glaeser, E. L. (Ed.), Agglomeration Economics. University of Chicago Press, Ch. 4, pp. 133–150.

Parent, O., LeSage, J. P., 2012. Spatial dynamic panel data models with random effects. Regional Science and Urban Economics 42 (4), 727–738.

Parr, J. B., Apr. 2002. Missing Elements in the Analysis of Agglomeration Economies. International Regional Science Review 25 (2), 151–168.

Perroux, F., 1950. Economic Space: Theory and Applications. The Quarterly Journal of Economics 64 (1), 89–104.

Perroux, F., 1955. Note sur la notion de pôle de croissance. Economie Appliquée 8, 307–320.

Porter, M. E., 1990. Competitive Advantage of Nations. Free Press.

Porter, M. E., 1998. Clusters and competition. On competition 76 (6), 197–287.

Puga, D., Venables, A. J., 1996. The Spread of Industry: Spatial Agglomeration in Economic Development.

Rey, S. J., 1998. The Performance of Alternative Integration Strategies for Combining Regional Econometric and Input-Output Models. International Regional Science Review 21 (1), 1–35.

Rey, S. J., 2000. Integrated regional econometric+input-output modeling: Issues and opportunities. Papers in Regional Science 79 (3), 271–292.

Rey, S. J., Dev, B., 1997. Integrating Econometric and Input-Output Models in a Multiregional Context. Growth and Change 28 (2), 222–243.

Rey, S. J., Jackson, R. W., 1999. Labor-productivity changes in regional econometric+ input-output models. Environment and Planning A 31, 1583–1599.

Robert-Nicoud, F., Jun. 2006. Agglomeration and Trade with InputOutput Linkages and Capital Mobility. Spatial Economic Analysis 1 (1), 101–126.

Rosenthal, S. S., Strange, W. C., 2001. The Determinants of Agglomeration. Journal of Urban Economics 50 (2), 191–229.

Starrett, D., 1978. Market allocations of location choice in a model with free mobility. Journal of Economic Theory 17 (1), 21 – 37.

Venables, A. J., 1996. Equilibrium locations of vertically linked industries. International Economic Review 37 (2), 341–359.

von Thünen, J. H., 1826. Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie, Hamburg. English translation by C.M. Wartenberg: von Thünen's Isolated State, Oxford, Pergammon Press (1966), Hamburg, Perthes.

Weber, A., 1909. Ueber den Standort der Industrien. J.C.B. Mohr, Tübingen, Germany.